



**MARIA CAROLINA
PEIXOTO ANTUNES**

**ESTUDO POPULACIONAL DE 10 *LOC*/X-STR NUMA
AMOSTRA DA POPULAÇÃO DE TIMOR-LESTE.**



**MARIA CAROLINA
PEIXOTO ANTUNES**

**ESTUDO POPULACIONAL DE 10 *LOC*/X-STR NUMA
AMOSTRA DA POPULAÇÃO DE TIMOR-LESTE.**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica do Professor Doutor Luís Souto Miranda, Professor Auxiliar Convidado do Departamento de Biologia da Universidade de Aveiro

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrônicos, quer de trabalhos académicos.

Dedico este trabalho aos meus filhos Rui e Ana para incentivar o seu estudo ao longo das suas vidas.

Aos meus pais Lina e Zé por todo o sacrifício de uma vida inteira e investimento que fizeram na minha formação.

À minha tia Ló agradeço a correção do texto.

Quero ainda deixar uma palavra de apreço e gratidão ao meu tio João por ter acreditado em mim e nas minhas capacidades.

Obrigada a todos pelo apoio.

o júri

presidente

Professora Doutora Maria de Lourdes Gomes Pereira

Professora Associada, com Agregação, do Departamento de Biologia da Universidade do Aveiro

Doutora Sandra Isabel Moreira Pinto Vieira

Professora Auxiliar Convidada da Secção Autónoma de Ciências da Saúde da Universidade de Aveiro

Doutor Luís Manuel Souto de Miranda

Professor Auxiliar Convidado do Departamento de Biologia da Universidade do Aveiro

agradecimentos

Ao Professor Doutor Luis Manuel Souto de Miranda, responsável pelo Laboratório de Genética Aplicada, do Departamento de Biologia da Universidade de Aveiro, quero agradecer por me ter recebido no seu laboratório, pelo desafio que me colocou ao propor o tema sobre o qual disserto - desafio que abracei de imediato por corresponder às minhas expectativas e vontade de alargar os meus conhecimentos académicos e profissionais. Mas sobretudo quero reconhecer a eficácia da orientação tendo em conta as diversas dificuldades que surgiram, nomeadamente, as relacionadas com meu estatuto de aluna trabalhadora, em conjunto com as vicissitudes que um trabalho prático desta natureza acarreta. Nos momentos em que precisei dispensou-me a atenção necessária e teve sempre uma palavra de orientação.

Agradeço também, à Mestre Filipa Tavares, mestre em Ciências Forenses pela Faculdade de Medicina da Universidade do Porto, à Mestre Helena Moreira, mestre em Medicina Legal pelo Instituto de Biomédicas de Abel Salazar da Universidade do Porto, à Mestre Helena Costa, Mestre em Biologia Molecular e Celular pela Universidade de Aveiro e à Doutora Sybelle Gomes, doutorada pela Universidade de Aveiro, colaboradoras no Laboratório de Genética Aplicada, do Departamento de Biologia da Universidade de Aveiro, com as quais aprendi muito, além de terem prestado um inapreciável apoio na execução prática deste trabalho, são exímias nas boas práticas de laboratório, pelo que lhes estou muito grata.

Ao Doutor Rui Pereira, colaborador no IPATIMUP, agradeço por tornar tão familiar e intuitivo o software Arlequin®, a principal ferramenta de tratamento estatístico de dados em genética populacional.

À Doutora Carla Santos, colaboradora no Grupo de Genética da Universidade Santiago de Compostela, pela simpatia, pelas sugestões e apoio, pois conhecemo-nos mesmo no início deste trabalho, quando eu era apenas uma curiosa da genética. As suas palavras, alegria e motivação inspiraram-me muito.

Ao Doutor Sizenando Abreu, a desempenhar funções no CESAM, pela sua amizade, companhia, pela troca de ideias e de conhecimento que tantas vezes me permitiram desbloquear o pensamento, evoluindo para o passo seguinte e transmitindo ânimo nos momentos mais difíceis.

Finalmente, quero agradecer ao Doutor Ângelo Ferreira, Técnico Superior na Reitoria da Universidade Aveiro, o desenho utilizado para ilustrar uma das páginas desta dissertação. Ângelo Ferreira esteve envolvido em várias atividades de desenvolvimento em Timor Leste, por exemplo o Projeto de Reestruturação do Ensino Secundário Geral, motivo que o levou a residir em Timor durante largos períodos de tempo, tendo criado laços profundos com os habitantes e a cultura timorenses deixando alguns registos de desenhos, entre outros legados.

palavras-chave

STR, Cromossoma X, PCR, X-STR GEP-ISFG Decaplex, Investigator Argus X-12; Timor Leste, genética populacional, genética forense, X-Decaplex, Argus X-12.

resumo

O presente trabalho propõe-se caracterizar uma amostra da população de Timor Leste por genotipagem com o “X-STR Decaplex do GHEP-ISFG”, um multiplex que amplifica por PCR 10 *loci* do cromossoma X (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 e DXS6789). Para o efeito utilizou-se uma amostra composta por 40 indivíduos, independentes entre si, habitantes de Timor Leste, país localizado no sudeste asiático. As frequências alélicas e os parâmetros de interesse forense para cada X-STR foram calculados e analisados por recurso a software Arlequin suite ver 3.5 e à calculadora do sítio Forensic ChrX Research, o sítio internacional do cromossoma X.

Principais resultados decaplex

Os dados referidos foram comparados com os obtidos por genotipagem de uma amostra de 134 indivíduos, também de Timor Leste, com o multiplex “Investigator Argus X-12”, que estuda 12 *loci* do cromossoma X (DXS8378, DXS7132, DXS7423, DXS10103, DXS10134, DXS10074, DXS10101, DXS10135 e DXS10146, DXS10079, HPRTB, DXS 10148).

As frequências alélicas e os parâmetros de interesse forense para cada X-STR foram calculados e organizados em formato de tabela.

Todos os marcadores estudados apresentam diversidades genéticas elevadas, únicas em relação a outras amostras populacionais referidas em literatura.

Comparação estatística dos dois métodos de caracterização; X-STR Decaplex GHEP-ISFG vs Investigator Argus X-12.

keywords

STR, X Chromosome, PCR, X-STR GEP-ISFG Decaplex, Investigator Argus X-12; East Timor, population genetics, forensic genetics.

abstract

This study aims to characterize a sample of 40 independent individuals, from East Timor, a country geographically placed in Southeast Asia, with X-STR Decaplex GHEP-ISFG (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789) by polymerase chain reaction.

All the markers studied show significant genetic diversity similar to those described in the literature for other population data.

Allele frequencies and forensic parameters of interest for each X-STR were calculated by Arlequin suite ver 3.5 and with the online calculator at Forensic ChrX Research internet site, then were organized in table format.

The procedure was also applied to a sample of 134 independent individuals, from East Timor using multiplex Investigator Argus X-12 (DXS8378, DXS7132, DXS7423, DXS10103, DXS10134, DXS10074, DXS10101, DXS10135 and DXS10146, DXS10079, HPRTB, DXS 10148).

All studied markers have high genetic diversity when compared to other population data reported in the literature. Allele frequencies and parameters of forensic interest for each X-STR were calculated and organized in table format.

Statistical comparison of the two methods of characterization was applied to X-STR-Decaplex-GHEP and Investigator Argus X-12

<< ...a ilha de Timor onde nasce o sândalo...>>



Desenho da autoria de Ângelo Ferreira.

“Vontade de bem fazer.”

Tradução a partir da divisa “Talant de bien faire”, visível no túmulo do
Infante Dom Henrique de Avis
ou
Infante de Sagres, *O Navegador*.

ÍNDICE

Índice de Figuras	iii
Índice de Tabelas	v
Abreviaturas.....	vii
Capítulo I Introdução.....	1
1 <i>O Genoma Humano</i>	2
1.1 O Projeto do Genoma Humano.....	5
1.2 Polimorfismos.....	7
1.3 Gene, Alelo e STR.....	8
1.4 Uso de STRs como Marcadores Genéticos.....	10
1.5 Genética Populacional, Perfil Genético e Identificação Humana.....	11
1.6 Princípio de Hardy-Weinberg.....	13
1.7 Cromossoma X.....	16
1.8 Modo de Transmissão do Cromossoma X & Compensação da Dosagem Genética.....	21
1.8.1 Lei de Lyon - Inativação do Cromossoma X ou Lyonização.....	22
1.8.2 Mecanismo de Inativação do Cromossoma X.....	24
1.9 Polimorfismos do Cromossoma X vs Genética Populacional.....	26
1.9.1 Polimorfismos de cromossoma X em contexto forense.....	28
1.10 Estudos populacionais na região do sudeste asiático.....	28
Capítulo III Objetivos.....	33
2 <i>Enquadramento</i>	34
2.1 Objetivo Geral.....	35
2.1.1 Objetivo Específico.....	35
Capítulo IV Material & Métodos.....	37
3 <i>Amostragem</i>	38
3.1 Extração e Quantificação de DNA genómico	38
3.2 Amplificação X-STR Decaplex.....	40
3.3 Amplificação <i>Investigator®</i> Argus X-12.....	44
3.4 Eletroforese em Gel de Agarose - AGE.....	45
3.5 Eletroforese Capilar e Genotipagem.....	47

Capítulo V	Resultados.	51
4	<i>Análise dos marcadores X-STRs</i>	52
4.1	Diversidade Genética	53
4.2	Comparação de Métodos	58
4.3	Estatística populacional para efeitos forenses.	59
Capítulo VI	Conclusão.	61
5	<i>Discussão e Conclusão.</i>	62
Capítulo VII	Bibliografia.	67

Índice de Figuras

Figura 1: Purinas e pirimidinas constituintes do DNA.....	3
Figura 2: Pentose: β -D-ribose.	3
Figura 3: Nucleosídeos constituintes do DNA.	4
Figura 4: Síntese de nucleotídeos, exemplo da adenosinamono-fosfato.	4
Figura 5: Cadeia de DNA.	5
Figura 6: Descrição da composição do DNA nuclear.	7
Figura 7: Short Tandem Repeat, STR.....	9
Figura 8: Perfil homozigótico. Herda cópia igual dos pais, para um dado locus, cujo STR contém 6 repetições da sequência repetitiva, um tetrâmero AATG. Logo, herda dois alelos 6 que originam um perfil 6/6.	9
Figura 9: Perfil heterozigótico, herda cópias diferentes, um alelo 6 da mãe e um alelo 4 do pai, logo o perfil é 6/4.	9
Figura 10: Diferentes Sistemas de Determinação Sexual.	18
Figura 11: Cariótipo humano masculino. As setas indicam o cromossoma X e Y.	19
Figura 12: Modelo de diferenciação dos cromossomas XY a partir do par dos autossomas ancestrais.	20
Figura 13: Mapa das camadas evolutivas nos cromossomas X e Y.....	20
Figura 14: Representação ideográfica das regiões PAR1 e PAR2 no par de cromossomas sexuais.....	21
Figura 15: Ciclo de inativação e reativação do CrX.....	25
Figura 16: Ideograma do Cromossoma X. Mapa dos 54 loci STR estudados para fins forenses.	27
Figura 17: Mapa das primeiras migrações humanas em direção ao sudeste asiático.	29
Figura 18: Movimentos migratórios que contribuíram para a pool genética de Timor-Leste.....	30
Figura 19: Marcadores do X-STR Decaplex do GHEP-ISFG.	43
Figura 20: Marcadores do Investigator® Argus X-12.	44
Figura 21: Representação do procedimento de separação do produto amplificado por eletroforese em gel de agarose, AGE.	46
Figura 22: Exemplo de uma eletroforese em gel de agarose obtido após genotipagem com o X-STR do Decaplex GHEP_ISFG.	46

Índice de Figuras

Figura 23: Esquema ilustrativo de um equipamento de Eletroforese capilar.....	47
Figura 24: Eletroferograma do perfil genético para o padrão 9947A com o painel X-STR Decaplex do GHEP-ISFG.	48

Índice de Tabelas

Tabela 1: Estudos genéticos disponíveis sobre amostras da população de Timor-Leste.....	34
Tabela 2: Tabela de Primers para o X-STR Decaplex GHEP-ISFG, Leonor Gusmão et al., 2009.....	41
Tabela 3: Preparação das misturas reacionais.	42
Tabela 4: Ciclo de amplificação para PCR.....	42
Tabela 5: Caracterização dos fragmentos de amplificação de acordo com o Sistema de primers do X-STR Decaplex do GHEP-ISFG	52
Tabela 6: Frequências alélicas e valores de diversidade genética para os 10 loci STR do X-STR Decaplex-ISFG, numa amostra de Timor-Leste.....	53
Tabela 7: Frequências alélicas e valores de diversidade genética para os 12 loci STR do Investigator® Argus X-12, numa amostra de Timor-Leste.....	54
Tabela 8: Comparação dos valores obtidos para os loci comuns nos multiplexs em estudo, DXS8378, DXS7132 e DXS7423.....	55
Tabela 9: Comparação Heterozigosidade ou diversidade genética para os 3 STRs comuns.	56
Tabela 10: Comparação do PIC para os 3 STRs comuns.....	56
Tabela 11: Comparação do PD female para os 3 STRs comuns.	57
Tabela 12: Comparação do PDmale para os 3 STRs comuns.	57
Tabela 13: Comparação entre H_{ESP} obtidos para cada multiplex.	58
Tabela 14: Comparação entre a Diversidade Genética obtida para o Argus X-12 e X-Decaplex	58
Tabela 15: Parâmetros de estatística populacional com interesse forense para 10 loci STR numa amostra de Timor-Leste. (n= 45 cromossomas X; 40 indivíduos)	60
Tabela 16: Parâmetros de estatística populacional com interesse forense para os 12 loci STR do Investigator® Argus X-12, numa amostra de Timor-Leste.	60
Tabela 17: Comparação entre populações para resultados obtidos após estudo com os 10 loci do X-STR Decaplex-ISFG-PD _{female}	63

Abreviaturas

AGE	Agarose Electrophoresis, eletroforese em gel de agarose	Kya	Thousand years ago, unidade de tempo utilizada na astronomia, geologia e biologia
ADN	Ácido desoxirribonucleico	M	indivíduos de sexo masculino
AS	Autossomas e autossómicos	Mb	Mega-basepairs, mega pares de bases
bp	base pair, o mesmo que pb, par de bases	Mya	Million years ago
CrX	Cromossoma X	PIC	Conteúdo informativo polimórfico
CrY	Cromossoma Y	PD	Poder de discriminação
ddp	diferença de potencial, volt	PDf	Poder de discriminação em mulheres
DNA	Desoxirribonucleic acid, sigla em inglês para ADN	PDm	Poder de discriminação em homens
DOE-NIH US	Department of Energy - National Institutes of Health United States	Pág.	Página
F	indivíduo sexo feminino	PAGE	PolyAcrylamide Gel Electrophoresis, eletroforese em gel de policacrilamida.
FTA®	Flinders Technology Associates, marca registada.	PAR1	Pseudoautosomal Region 1
GHEP-ISFG	Grupo de Habla Española y Portuguesa de la ISFG / Grupo de Línguas Espanhola e Portuguesa da ISFG	PCR	Polymerase Chain Reaction
GEP-ISFG	Grupo Español y Portugués de la ISFG, anterior designação do GHEP-ISFG	pb	par de bases
H-W	Hardy-Weinberg	SDS	Sodium Dodecyl Sulfate ou dodecil-sulfato de sódio, em português.
ISFG	International Society of Forensic Genetics	SNV	Single Nucleotide Variants
IUBMB	International Union of Biochemistry and Molecular Biology	STR	Short Tandem Repeats
IUPAC	International Union for Pure and Applied Chemistry	STS	Short Tandem Sequences
		SSR	Simple Sequence Repeat
		SRY	Sex-determining Region Y
		TBE	Tris-borato EDTA
		VNTR	Variable Number Tandem Repeat
		XAR	X chromosome Added Region
		XCR	X chromosome Conserved Region

*“Não é a espécie mais forte que sobrevive,
ou a mais inteligente,
mas a que melhor se adapta à mudança.”*
Charles Darwin

Capítulo I Introdução

1 O Genoma Humano.

O genoma humano é todo o ADN existente num organismo humano.

O genoma humano de referência é composto pelo ADN celular, presente no núcleo da célula, corresponde a 23 pares de cromossomas, dos quais 22 pares são autossomas (numerados de 1 a 22) mais 1 par composto pelos cromossomas sexuais (XY), e por uma molécula de ADN mitocondrial (presente exclusivamente nas mitocôndrias espalhadas pelo citoplasma). Os cromossomas correspondem ao ADN nuclear, cada um é constituído por uma cadeia de ADN linear e aberto nas extremidades, mas enrolado sobre si mesmo, tendo por base da estrutura os nucleossomas. O ADN mitocondrial é constituído por uma cadeia de ADN circular, do qual uma parte é designada por cadeia pesada.

Todas as células contêm, virtualmente, as mesmas cópias de ADN, perfazendo um total de 46 cromossomas que, por sua vez, são constituídos por ácido desoxirribonucleico de cadeia dupla, em espiral e aberto nas extremidades, e, por moléculas de ácido desoxirribonucleico mitocondrial (presente exclusivamente nas mitocôndrias e igual entre si) de cadeia dupla, fechada e em forma circular. Exceção feita para as células haplóides ou gametas, compostas por uma unidade de cada cromossoma em vez de pares de cromossomas, tendo metade do ADN das células somáticas, ou células diplóides.

O genoma (número total de cromossomas constituído por ADN codificante ou genes e por ADN não codificante) e o epigenoma (diferentes marcas colocadas ao longo dos cromossomas, por exemplo, metilação dos grupos CpG do ADN, uso de diferentes tipos de histonas na formação dos nucleossomas, e sua metilação, acetilação, ou ligação de outros grupos químicos às histonas, que permitem ativar ou silenciar a informação contida no ADN) (UCL, 2015) têm a informação necessária para construir uma nova célula (Trerotola et al, 2015).

O ácido desoxirribonucleico ou ADN é regularmente designado pelas siglas inglesas DNA conforme a nomenclatura aprovada em conjunto pela Sociedade Internacional de Química Pura e Aplicada (IUPAC) e pela Sociedade Internacional

de Bioquímica e Biologia Molecular (IUBMB) (Biochemical Nomenclature Committees, 1974). Assim será baseada toda a terminologia utilizada nesta dissertação exceto quando devidamente fundamentada. Esta nomenclatura foi adoptada por forma a normalizar e facilitar a comunicação oral e escrita entre os grupos científicos desta área espalhados por todo mundo.

O DNA é um polímero cuja unidade base são os nucleotídeos, cada nucleotídeo é constituído por uma base purínica ou pirimidínica, uma pentose e um fosfato (Dose, 1982). As bases púrinicas, ou purinas, são a adenina e a guanina e as bases pirimidínicas, ou pirimidinas, são a citosina e a timina (Figura 2). A pentose é a β -D-ribose (Figura 2), também designada por ribofuranose.

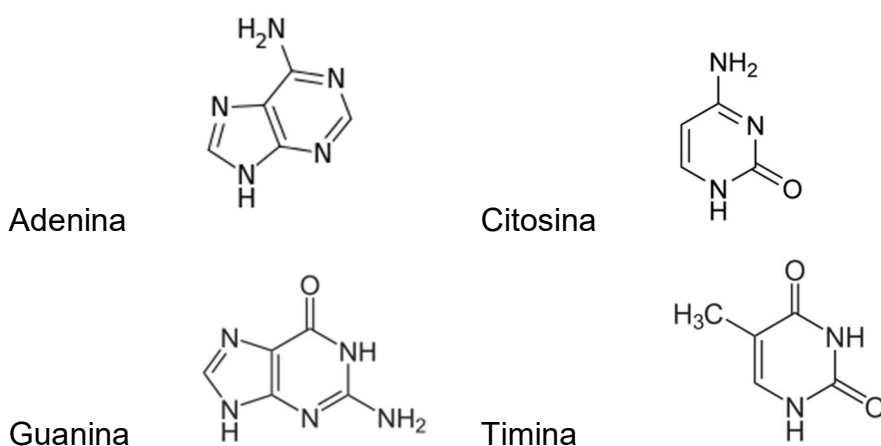


Figura 1: Purinas e pirimidinas constituintes do DNA.

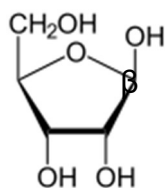


Figura 2: Pentose: β -D-ribose.

A uma purina ou pirimidina ligada a uma β -D-ribose constitui um nucleosídeo. Os nucleosídeos que constituem o DNA (Figura 3) são a adenosina, Ado ou na forma de radical pela letra A; guanossina, Guo ou na forma de radical por G; citidina,

Cyd ou na forma de radical por C e timidina, Thd ou na forma de radical por T (Biochemical Nomenclature Committees, 1974).

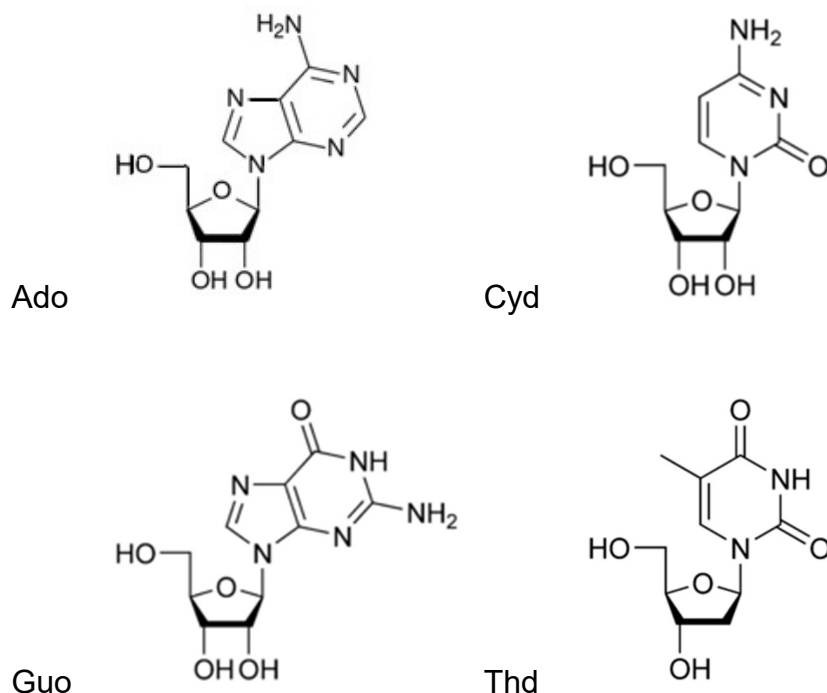


Figura 3: Nucleosídeos constituintes do DNA.

O grupo fosfato liga-se na posição 5' da ribose presente no nucleosídeo (Carey, 1992):

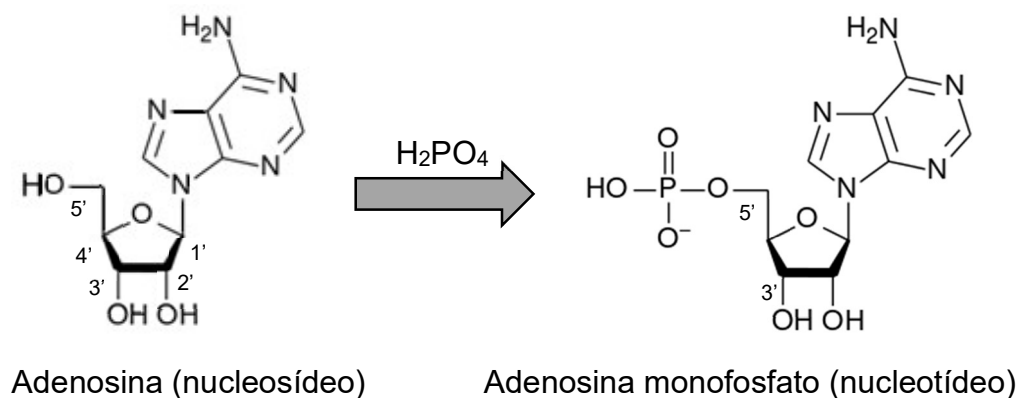


Figura 4: Síntese de nucleotídeos, exemplo da adenosinamofosfato.

O fosfato do nucleótido seguinte liga-se na posição 3' da ribose e assim sucessivamente (Figura 4), formando uma cadeia simples de DNA, no sentido de 5' para 3', sendo a presença do fosfato que confere carga negativa à cadeia polimérica.

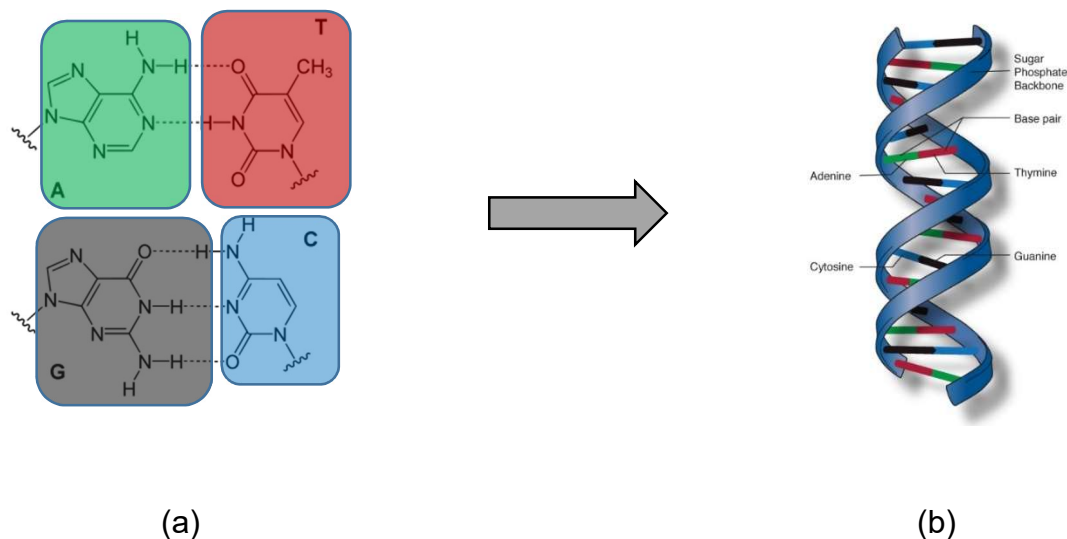


Figura 5: Cadeia de DNA.

(a) Estrutura das bases azotadas (purinas e pirimidinas) emparelhadas por pontes de hidrogénio, isto é, apresentação dos pares de bases.

(b) Estrutura DNA em dupla hélix.

Adaptado de Wikimedia Commons, 2014, e de Darryl Leja, 2010, (a) e (b) respetivamente.

Nota: As cores na figura (a) correspondem às mesmas bases representadas na figura (b).

A cadeia em dupla hélix forma-se quando duas cadeias de DNA se ligam por pontes de hidrogénio entre os respetivos pares de bases purina/pirimidina.

1.1 O Projeto do Genoma Humano.

O Projeto Genoma Humano, iniciado em 1990, foi coordenado, durante 13 anos, pelo Departamento de Energia e vários Institutos Nacionais de Saúde dos Estados Unidos (DOE-NIH US).

O DOE procede da união entre as anteriores Comissão da Energia Atómica e a Administração para a Investigação e Desenvolvimento da Energia nos Estados Unidos, dado estas entidades terem sido as responsáveis pelo desenvolvimento da bomba atómica e, após o seu uso, que culminou com o fim da Segunda Guerra Mundial, o congresso americano responsabilizou-os pelos seguintes estudos: análise da estrutura do genoma e sua replicação; se o dano cometido pelas radiações no DNA seria possível de reparação; quais as consequências das mutações genéticas, especialmente aquelas causadas pela radiação e pelos subprodutos químicos originados a longo prazo pelas explosões atómicas, com o objetivo de diagnosticar e tratar as vítimas causadas pela sua ação e as consequências nos seus descendentes.

A partir dos estudos desenvolvidos constatou-se que a melhor maneira para estudar os efeitos seria analisando todo o genoma humano de forma a obter uma sequência de referência. À medida que os trabalhos se foram aprofundando verificou-se que as tarefas teriam que ser levadas a cabo em conjunto com especialistas da área da saúde e especialistas noutras áreas da ciência, em particular com químicos, biólogos, matemáticos e informáticos, por terem a formação e os recursos adequados para concretizar o objetivo (DOE-US, 2013 e DOE-US, 2006).

A DOE começou o planeamento do Programa Genoma Humano em 1986 e em 1987 iniciou o planeamento conjunto com o Programa de Saúde Americano, envolvendo desta forma os Institutos Nacionais de Saúde dos Estados Unidos (NIH). O Projeto Genoma Humano teve início formal a 1 de outubro de 1990, após 5 anos de reuniões tendo-se elaborado o primeiro plano conjunto e assinado um memorando de entendimento entre as duas organizações.

O projeto foi planeado para ter uma duração de aproximadamente 15 anos, mas o trabalho árduo e o desenvolvimento da tecnologia aceleraram o seu fim para 2003. Duas das principais metas do Projeto Genoma Humano eram identificar todos os genes humanos e determinar o número de pares de bases que compõem uma sequência completa do genoma humano, assim sendo, identificaram-se mais de 20 000 genes codificadores de proteínas e RNA, e perto de 3×10^9 de pares de bases, ou 3234,83 Mb, como constituintes do DNA nuclear de uma célula diploide ("Human Genome Project Information Archive - About the Human Genome Project," 2014). O DNA codificante representa aproximadamente 3% do genoma e o DNA não codificante representa os restantes 97% (Bennett, 2000) (Figura 6).

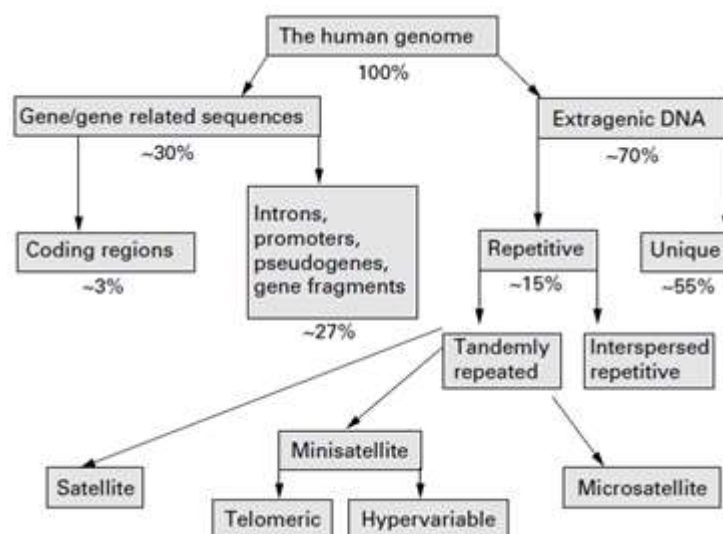


Figura 6: Descrição da composição do DNA nuclear.
Adaptado de (Bennett, 2000)

Os *Homo sapiens* são iguais em 99,7% do seu genoma restando 0,3% do genoma que permita distinguir os seres humanos uns dos outros.

1.2 Polimorfismos.

Os marcadores genéticos utilizados em estudos de genética populacional e na genética forense estão localizados nas regiões não codificantes incluindo intrões; não se recorre às regiões codificantes de proteínas nem às regiões reguladoras da expressão de proteínas por razões de ética, esta restrição não é limitativa pois, como se sabe, esta porção corresponde apenas a 3% do total.

Existem imensas sequências de pb que se repetem, espalhadas pelo DNA que presentes em todos os indivíduos, mas variam na forma da sequência repetitiva, na sua versão e na sua expressão, de indivíduo para indivíduo e como tal são designadas por polimorfismos do DNA. Cada zona onde se encontram estes polimorfismos é designada por *locus*, palavra latina que significa localização, sendo *loci* o seu plural.

As repetições longas podem conter entre centenas a vários milhares de pb em repetição e designam-se por minissatélite ou VNTR.

As repetições mais pequenas contêm unidades de repetição que variam de 2 a 7 pb e, por serem mais específicas, passaram a designar-se por microsatélites, STS,

SSR ou mais vulgarmente STR. A maioria dos STRs encontram-se localizados nas regiões não codificantes (Li et al, 2002).

A designação DNA satélite teve origem em estudos de gradiente de densidade por centrifugação, realizados em 1960. Os investigadores observaram um pico de DNA que surgia constante e sistematicamente separado do restante e por isso o designaram DNA satélite. Essa porção de DNA separado foi estudada e verificaram que consistia em sequências de nucleótidos base, consecutivamente repetidas, constituindo um conjunto, isto é, em *tandem* (Ellegren, 2004).

Os SNP são unidades de repetição de um pb apenas e tornaram-se a ferramenta mais precisa e detalhada em estudos de genética populacional em conjunto com marcadores INDEL (*insertion/deletion* que permitiu estudo das mutações para os SNP).

1.3 Gene, Alelo e STR.

O gene é um segmento de DNA cuja sequência de nucleótidos contém o código com as instruções para que as estruturas celulares construam as proteínas e RNA necessários para a estrutura dos organelos ou para desempenhar outras funções na célula.

O alelo é a versão de um gene para um dado *locus*, este pode ter uma ou mais versões, isto é, produzindo proteínas diferentes, gerando diferentes fenótipos, por exemplo: o gene que origina o grupo sanguíneo baseado no Sistema ABO, descoberto por Karl Landsteiner, tem 3 alelos A, B e O, que posteriormente foi designado pelo número 0 (zero), que produzem 4 fenótipos (A, AB, B e O), ao mesmo tempo que o gene que produz o Sistema baseado na presença do Fator Rh tem apenas duas versões, dois alelos, R e r, produzindo o fenótipos R⁺ (a partir dos genótipos RR ou Rr) e R⁻ (genótipo rr). O genótipo é o conjunto de genes de cada indivíduo portadores de informação, herdados dos pais. E o fenótipo é a expressão de apenas alguns dos genes que constituem o genótipo, o fenótipo é o resultado da expressão dos genes ativos no indivíduo, as quais resultam nas características observáveis.

Normalmente, todos herdamos, uma única versão de cada polimorfismo, cada uma proveniente de cada progenitor. Se herdarmos duas versões iguais do mesmo

alelo, o perfil designa-se homozigótico (Figura 8), se herdarmos versões diferentes do mesmo alelo, o perfil designa-se heterozigoto (Figura 9); podemos herdar diferentes versões entre indivíduos e é aqui que reside a variabilidade entre os organismos, assim como, a possibilidade de os distinguir, identificando-os. Não existem dois indivíduos com os mesmos polimorfismos, exceto os gémeos monozigóticos ou verdadeiros e mesmo estes têm algumas marcas epigenéticas distintas, sendo este o novo passo a explorar na identificação de indivíduos.

O termo alelo também se aplica de igual forma para as zonas não codificantes, sendo que a versão do alelo corresponde ao número de vezes que se repete a unidade sequencial base para um dado locus, neste caso o alelo é designado por um número.

Os STRs (Figura 7) são segmentos que contêm sequências de nucleótidos que se repetem em *tandem*, isto é, em conjunto, e podem ser constituídos por repetições de 2 pb a 6 pb consecutivos ou intercalados por nucleótidos que se repetem de igual forma entre indivíduos, que correspondem a 3% do DNA (Dashnow et al., 2015; Lander et al., 2001).



Figura 7: Short Tandem Repeat, STR.

Cujo motivo de repetição base é o tetrâmero AATG e alelo é o 6 por ser o número de vezes que se repete na sequência de DNA.



Figura 8: Perfil homozigótico. Herda cópia igual dos pais, para um dado locus, cujo STR contém 6 repetições da sequência repetitiva, um tetrâmero AATG. Logo, herda dois alelos 6 que originam um perfil 6/6.

Figura 9: Perfil heterozigótico, herda cópias diferentes, um alelo 6 da mãe e um alelo 4 do pai, logo o perfil é 6/4.

1.4 Uso de STRs como Marcadores Genéticos.

Define-se como marcador genético toda a característica do genoma que se repita e que possa ser seguida de forma sistemática recorrendo a métodos de análise biomolecular adequados para o efeito.

Os STRs são os marcadores genéticos mais utilizados em estudos de genética populacional pelos seguintes motivos (Butler, 2012; Silva, 2007):

- Facilmente amplificados por PCR (polymerase chain reaction), permitindo recuperar informação a partir de amostras de DNA em baixa quantidade ou mesmo degradadas;
- Os STRs são suficientemente pequenos, o que permite que sejam independentes de outros locus, por isso, menos sujeitos ao efeito do desequilíbrio de ligação, menos sujeitos à formação de alelos *drop-out*, isto é, desaparecimento do alelo por falha na amplificação deste ou aparecer fora da zona eletroforética esperada, ou seja, os STRs estão menos sujeitos ao efeito alelo nulo por dificuldade de emparelhamento do *primer* no DNA;
- Ubiquidade, grande variabilidade de locus ao longo de todo o genoma compostos por STRs com elevado polimorfismo, apesar da porção de DNA reservada a este tipo de polimorfismo ser de 3% do genoma humano (Dashnow et al., 2015; Lander et al., 2001);
- Distribuição aleatória ao longo de todo o genoma;
- Reduzida formação de produtos *stutter* quando comparados com os obtidos por outro tipos de polimorfismos, o que facilita a interpretação de amostras com misturas de genótipos.
- Maior frequência de heterozigóticos e menor taxa de mutação.

Nomenclatura de um marcador STR para um dado *locus* segue a seguinte configuração que se exemplifica para o marcador D3S4529 (Kobachuk, 2012):

“D” de DNA;

“3” é o número do cromossoma a que pertence o *locus*;

“S”, de “sequência de cópia única” (do inglês *single copy sequence*)

“4529” o número do *locus* em que o marcador foi descrito.

1.5 Genética Populacional, Perfil Genético e Identificação Humana.

A genética populacional é a ciência que estuda a história e os mecanismos que originam a variação genética nas populações de espécies, revelando as relações entre populações e a evolução a partir de populações ancestrais. A população é constituída por um grupo de indivíduos da mesma espécie que potencialmente se cruzam (acasalam) entre si, os quais fornecem o seu DNA para estudo. Entre humanos a colheita é feita após consentimento informado.

Os estudos genéticos assentam na diversidade alélica, e, no número de vezes que esses alelos surgem numa dada população, parâmetro designado por frequência alélica.

O estudo do genoma humano permitiu tomar consciência desta variabilidade genética e de como, mesmo sendo pouco reduzida quando comparada com a totalidade do genoma, essa variabilidade permite conclusões seguras e irrefutáveis sobre:

- Ancestralidade humana, permitindo fazer inferências sobre a evolução da espécie humana, determinando a sua origem;
- Ascendência de um indivíduo realizando estudos de linhagem que permitem, por exemplo, criar árvores genealógicas fidedignas ou simplesmente determinar graus de parentesco entre indivíduos, sobretudo para efeitos forenses;
- Identificação de indivíduos, pois a frequência de um dado perfil genético é calculada a partir das frequências alélicas de um dado conjunto de marcadores na população à qual pertence, permitindo aferir se esse perfil genético se distingue de forma única ou se é um perfil comum a vários indivíduos. A identificação é tanto melhor quando os marcadores são específicos para aquela população e quanto maior a sua diversidade (Brenner, 2015).
- Identificação de genes causadores de doenças e mecanismos de expressão genética permitindo realizar o diagnóstico de doenças genéticas, prevendo a sua expressão nas populações e desenvolver formas de terapia genética.

Os mecanismos que condicionam o processo de evolução são:

- Deriva genética provocada por variações na frequência alélica devido ao facto da probabilidade de herdar certos alelos ser independente e aleatória (segunda lei de Mendel ou Lei da Segregação Independente). A deriva genética não faz com que os indivíduos sejam mais adaptáveis mas sim, e apenas, permite que sejam diferentes entre si.
- Seleção natural, os mais adaptáveis sobrevivem. Os alelos que tornam o indivíduo mais adaptável e resistente à vida na Terra são por isso mesmo propagados, ao passo que os outros indivíduos morrem ou têm esperanças de vida mais reduzidas evitando a propagação através da descendência.
- Seleção sexual ou cruzamento não aleatório, existem características que tornam uns indivíduos mais apelativos sexualmente do que outros o que induz o caminho da evolução da espécie, porque os alelos dos pares mais apelativos vão surgir mais frequentemente na população (*gene pool*).
- Mutação ocorre, geralmente, durante a meiose na produção dos gametas podendo levar à origem de novos indivíduos com mutações incompatíveis com a vida ou a mutações benéficas que poderão propagar-se ao longo de várias gerações.
- Fluxo genético que ocorre quando novos indivíduos, com novos conjuntos de alelos, espalham os seus alelos por toda a população (primeira lei de Mendel ou Lei da Segregação), por ação dos movimentos migratórios, e, tal como com a deriva genética os seus efeitos são observados em pequenas populações. Fenómeno observado na população Timorense, pequena e muito variável devido aos vários efeitos de povoamento desde tempos ancestrais até à atualidade. Os indivíduos nativos são essencialmente de origem malaia que posteriormente se cruzaram com indivíduos de várias origens, nomeadamente portugueses, holandeses, chineses, polinésios e aborígenes australianos, não necessariamente por esta ordem.

1.6 Princípio de Hardy-Weinberg.

O Princípio de Hardy-Weinberg é o princípio base do qual partem todos os estudos em genética populacional, permite através de dados genotípicos prever as frequências alélicas para população ideal e o desenvolvimento de novos genótipos. O princípio prevê que, para uma população mendeliana infinita, as frequências alélicas e genotípicas são constantes ao longo do tempo, logo a população não evolui (Goudet, 1993).

Este princípio foi batizado com o nome dos dois cientistas, Godfrey Hardy (Hardy, 2003) e Wilhelm Weinberg (Weinberg, 1908), que estudaram cada um *per si*, o mesmo assunto utilizando os mesmos pressupostos e chegaram exatamente à mesma equação, a qual descreve como a genética mendeliana funciona à escala de uma população inteira. De acordo com esses pressupostos assumiu-se que nenhuma das forças anteriormente referidas influenciam a população permitindo-a atingir um equilíbrio estável tal que as frequências genéticas/alélicas mostram uma relação constante entre si ao longo do tempo, isto é, que a população modelo se encontra no estado de equilíbrio de Hardy-Weinberg (H-W), logo qualquer desequilíbrio dos valores esperados será atribuído aos mecanismos acima referidos o que irá permitir inferir sobre quais os fatores que o causaram.

Por hipótese, a população em equilíbrio H-W deve ter as seguintes características (pressupostos), isto é:

- Uma população estática, que não evolui, logo não há seleção natural;
- *Gene pool* constante, isto é, o número e a frequência de alelos mantém-se constante de geração em geração logo:
- Não há seleção sexual e o cruzamento é aleatório;
- Não há ocorrência de mutações;
- Requer um tamanho grande de população amostra por forma:
- Não se sentir o efeito de deriva genética, pois quanto mais pequena a população mais evidente se torna a diferença entre indivíduos;
- Não se observar o fluxo genético entre populações devido a fenómenos de migração.

A população base do equilíbrio de H-W é vulgarmente, designada por “população chata”, numa tradução à letra a partir do inglês, por apresentar uma estabilidade

genética constante ao longo do tempo. Os valores estatísticos teóricos ou esperados para ela calculados constituem a hipótese nula, H_0 . Esta é uma forma de utilizar o princípio da redução ao absurdo por modo a interpretar, de forma imparcial e significativa, os desvios aos valores obtidos para uma população real, permitindo concluir sobre a sua estrutura e sua origem.

É em relação aos valores esperados (valores teóricos) para a população em equilíbrio H-W, e que constituem a hipótese nula (H_0), que se inferem as conclusões relativas aos valores obtidos para população real, ou hipótese alternativa, H_1 .

Numa observação superficial, aquilo que se influi sobre as características genéticas dos indivíduos que compõem uma dada população, é aquilo que se observa e que pode ser medido por descrição da característica isto é, com base no fenótipo, e contagem do número de indivíduos que a contêm.

O genótipo é determinado por técnicas biomoleculares, como a tipagem por PCR, ou deduzidos a partir de modelos.

Supondo que só existem dois alelos, A e a, para uma dada característica, sendo a sua frequência alélica designada por p e q, a soma das frequências é igual a 1, pois a população não está sujeita a qualquer variação. Assim sendo, como cada indivíduo herda $\frac{1}{2}$ genótipo do pai : $\frac{1}{2}$ genótipo da mãe, o genótipo completo é composto por dois alelos para cada característica que apresente, logo a frequência genotípica num indivíduo é $(p+q)^2=1$, sendo que esta é a equação de H-W.

A equação de H-W desenvolvida é equivalente a:

$$p^2+2pq+q^2=1$$

onde p^2 é a frequência do genótipo Homozigótico, AA

q^2 é a frequência do genótipo Homozigótico, aa.

e

$2pq$ é a frequência do genótipo Heterozigótico, Aa.

Os estudos de genética populacional consistem na determinação dos vários alelos para uma dada característica e cálculo da sua frequência. O cálculo das frequências é muito simples (número de vezes que um alelo se repete, para um dado locus, a dividir pelo número total de cromossomas genotipados da amostra populacional).

As amostras são de tal modo grandes que o volume de dados obtidos levou ao desenvolvimento de ferramentas de cálculo para agilizar o tratamento de dados (existem dados obtidos a partir de mais de 1000 indivíduos, como é o caso de estudos em populações asiáticas), assim como, além do cálculo das frequências, existem outros parâmetros cujas equações são mais complexas e que se tornam mais rápidos de obter por recurso a diversos softwares, por exemplo, o cálculo da heterozigosidade observada (HETobs ou diversidade genética), heterozigosidade esperada (HETesp), percentagem de *loci* polimórficos, número médio de alelos por *locus*, distância genética, estatística F ou índices de fixação (F_{is} , F_{it} , F_{st}), fluxo génico, conteúdo de informação polimórfico (PIC), probabilidade de exclusão (PE), desequilíbrio de ligação (LD), poder de discriminação (PD) entre outros.

No caso de estudos que recorrem a genótipos obtidos com cromossoma X, os cálculos estatísticos que têm por base o estudo do desequilíbrio de Hardy-Weinberg só podem ser aplicados a perfis genéticos femininos, pois apenas as fêmeas recebem dois CrX, um alelo no CrX do pai e o outro do CrX da mãe, conforme se explicará mais à frente (Schaffner, 2004). Os perfis masculinos são hemizigóticos, uma vez que herdaram um CrX recombinado, herdado da mãe, permitem o estudo de linhagens, criação de haplótipos e o cálculo de LD.

O haplótipo é a designação para o conjunto de marcadores genéticos, neste caso STR do CrX, cuja localização no cromossoma se apresenta longe das habituais zonas de recombinação e a proximidade entre si permite que sejam transmitidos em conjunto.

O desequilíbrio de ligação, LD, é a associação de dependência, ou combinação não-aleatória, entre *loci* ou alelos (dois ou mais *loci*) localizados geralmente no mesmo cromossoma que resulta na formação de haplótipos.

1.7 Cromossoma X

Desde sempre que a mais conhecida forma de polimorfismo genético entre a maioria dos indivíduos da mesma espécie é o dimorfismo sexual (Zarkower, 2001). Ou se nasce com cariótipo masculino XY (Figura 11) ou se nasce XX, cariótipo feminino, sendo que as restantes formas humanas decorrem de erros genéticos nos gametas, durante a fecundação ou nos primeiros dias de gestação após a fecundação.

Apesar de atualmente sabermos que o género é herdado geneticamente, só muito recentemente se descobriu que essa herança se deve à atividade cromossómica, pois na primeira metade do século XX, a ideia de que o sexo poderia ser determinado pelos cromossomas era algo novo.

Inicialmente a determinação do sexo era atribuída a outros fatores, tais como o meio em que os ovos/zigotos estavam expostos no ventre da mãe, nomeadamente, temperatura e alimentação, de tal forma que ainda hoje nos recordarmos do tempo em que a responsabilidade do sexo da criança era apenas atribuída ao comportamento da mãe.

Hoje em dia, conhecemos bem o genoma humano contido no cariótipo composto por 23 pares de cromossomas; sabemos como se formam dos gametas através do processo de meiose e o fenómeno da fecundação; que todos os cromossomas existem em pares homólogos ou autossomas, exceto os cromossomas sexuais; que os cromossomas são numerados, exceto os cromossomas sexuais, os quais são designados pelas letras X e Y. Sabemos ainda que a determinação do sexo do indivíduo apenas depende da herança genética, as mulheres herdam XX e os homens XY.

Qual a origem desta terminologia para os cromossomas sexuais, qual é a sua história científica?

A origem da designação “cromossoma X” para o cromossoma feminino não é muito clara mas tornou-se tão familiar que raramente nos questionamos sobre como surgiu o seu nome (Balderman & Lichtman, 2011; Zimmer, 2014).

Hermann Henking (Berger, 2008) foi o primeiro a utilizar o termo “elemento X” para designar um dos cromossomas sexuais.

Foi em 1891 que pela primeira vez utilizou a designação “elemento X” para se referir a uma porção de cromatina presente no núcleo das células dos *Pyrrhocoris apterus* ou percevejo vermelho e preto.

Durante os seus projetos, Henking observou que durante a anáfase da segunda divisão meiótica, havia uma pequena porção de cromatina que parecia não estar em dobro e designou-a por elemento X (Henking, 1891). Este cromossoma movia-se para os polos, aparentemente sem sofrer divisão, deixando para trás os outros cromossomas, originando células gaméticas com 11 e 12 cromossomas, dando origem a fêmeas com 24 cromossomas (12 pares de cromossomas homólogos) e machos com 23 cromossomas (11 pares homólogos mais um cromossoma solitário, o elemento X - para Henking).

Em 1902, estudos elaborados por McClung atribuíram pela primeira vez a responsabilidade na determinação sexual, naqueles insetos, ao cariótipo herdado, e designou por cromossoma acessório ao cromossoma solitário (Allen, 2008b; McClung, 1901, 1902).

Não se sabe se Henking designou a porção de cromatina por “X”, como o fazem os matemáticos para designar uma incógnita numa equação, ou se por entender que era um cromossoma excluído durante o processo de meiose, um “ex-cromossoma”, sendo X abreviatura para o prefixo “ex-” de excluído, afastado, separado, isolado, em resumo, privado de par.

Observações semelhantes foram efetuadas por outros investigadores que só em 1903 (Wynbrandt & Ludman, 2009) identificaram indubitavelmente esta porção de cromatina como sendo um cromossoma sexual com características e modo de transmissão diferente entre espécies, mas com a mesma expressão fenotípica – a determinação do sexo.

Thomas Hunt Morgan (Allen, 2008a), inicialmente cético quanto à possibilidade da determinação do sexo ser um caractere hereditário, entendia que havia poucos estudos que o fundamentassem (Morgan, 1909), como tal, dedicou-se ao seu estudo e foi dos primeiros a identificar, sem sombra de dúvidas, os cromossomas X e Y no cariótipo da *Drosophila melanogaster* e revelou a sua função na determinação sexual daqueles organismos (Morgan, 1910).

Após sucessivas observações e vários anos de estudo, chegou-se à conclusão que o sexo dos organismos é determinado por diferentes mecanismos (Figura 10), entre os quais três têm por base sistemas cromossômicos (Carlson, 2013; Ilona Miko, 2008):

1. Sistema XX/XØ, onde Ø indica ausência, XX define as fêmeas e XØ os machos; ocorre sobretudo nos insetos;
2. Sistema XX/XY, onde XX define as fêmeas e XY os machos; ocorre na maioria dos mamíferos, em muitos dos artrópodes e algumas plantas.
3. Sistema ZZ/ZW, onde ZZ define os machos e ZW as fêmeas, ocorre nas aves, peixes, cobras, serpentes e vários insetos.

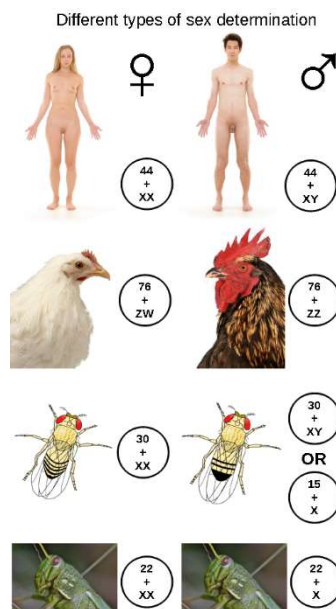


Figura 10: Diferentes Sistemas de Determinação Sexual.
Adaptado de Commons Wikipedia, "Sex-Determination System," 2014.

Nos seres humanos a determinação do sexo dos indivíduos é do tipo XX/XY.

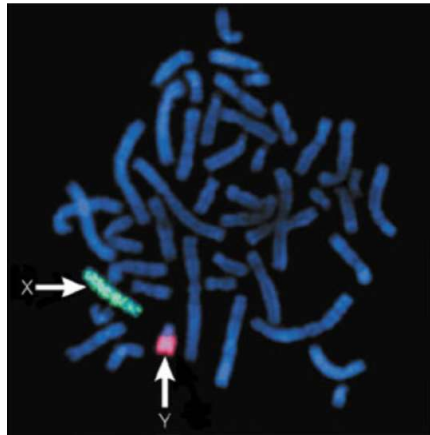


Figura 11: Cariótipo humano masculino. As setas indicam o cromossoma X e Y.
Imagem de Betsy Hirsch and Olga Shilko, adaptadas de Zarkower, 2001.

Os cromossomas X e Y humanos são morfologicamente e geneticamente muito distintos, mas existem evidências de que nem sempre terá sido assim e que terão evoluído a partir de um par de autossomas (Graves, 2010).

O cromossoma X é um cromossoma médio com 155 Mb e perto de 819 genes que codificam proteínas (Ensembl release, 2015a), é o cromossoma com menor densidade génica (Ross et al., 2005, 2009). Já o cromossoma Y é um dos três cromossomas mais pequenos com 59Mb e perto de 72 genes(Ensembl release, 2015b), sendo apenas precedido em tamanho pelos cromossomas 21 e 22, os seus genes têm funções praticamente específicas e exclusivas na formação dos caracteres masculinos que estão localizados na sua região SRY, Sex-determining Region Y.

O par de autossomas que originou os atuais cromossomas sexuais (Figura 12), provavelmente, terá começado a diferenciar-se num processo de evolução que consiste em quatro eventos ou etapas (Lahn & Page, 1999), assim designados conforme os autores. Durante cada etapa ocorre uma recombinação supressora XY num estrato ao longo do cromossoma Y, sem perturbar a ordem dos genes no cromossoma X.

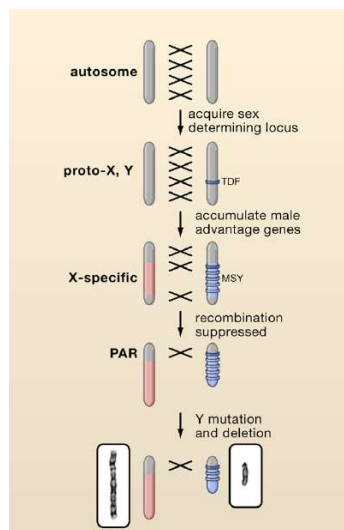


Figura 12: Modelo de diferenciação dos cromossomos XY a partir do par dos autossomos ancestrais. Adaptado de Graves, 2006.

O primeiro evento (Figura 13), que marcou o início de diferenciação XY, terá ocorrido há 240 - 320 milhões de anos atrás, logo após a divergência das linhagens de mamíferos e aves. (Lahn & Page, 1999).

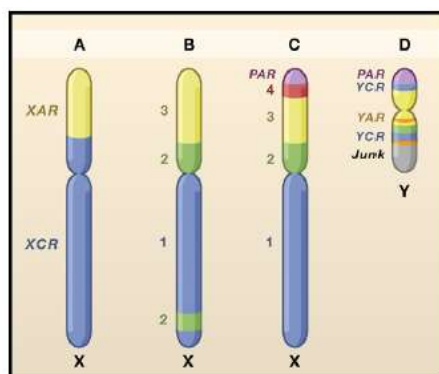


Figura 13: Mapa das camadas evolutivas nos cromossomos X e Y. Adaptado de Graves, 2006.

- (A) Mapa comparativo de genes de outros mamíferos presentes no CrX humano, realizado por ornitólogos. Na imagem observa-se uma região antiga conservada no CrX comum em todos os mamíferos (zona representada a azul, XCR) e uma região adicionada (zona representada a amarelo, XAR) que é autossômica nos mamíferos marsupiais e monotrêmatos que foi adicionada ao CrX 180-100 mya.
- (B) Genes da galinha homólogos aos do mapa do CrX humano para três regiões autossômicas por mapeamento comparativo. Esta análise identifica três estratos evolutivos que correspondem à região adicionada, XAR (camada 3) e subdivide em duas partes a região conservada, XCR (camadas 1 e 2).
- (C) A comparação entre a sequência de nucleótidos. Dos genes no CrY humano e seus homólogos no CrX permitiu a agrupá-los em 5 estratos. O estrato mais antigo corresponde na maioria ao estrato 1, a seguinte corresponde ao estrato 2. A zona XAR contém três porções a 3 e 4 e a região PAR.
- (D) Por meio de mapeamento comparativo direto entre os CrY humano e marsupial, e por analogia com os estratos evolutivos do X, podemos deduzir que o CrY humano, também, é subdividido em regiões antigas e adicionada. A antiga região conservada no CrY (azul, YCR) é minúscula, e a maior parte do Y deriva da região adicionada (YAR). Alguns genes (laranja) também foram transpostos de outros cromossomos e o restante é considerado lixo ou DNA não codificante.

Os cromossomas X e Y ainda hoje mantêm as zonas homólogas (Figura 14) e sofrem recombinação nas zonas designadas por regiões pseudoautossómicas (PAR).

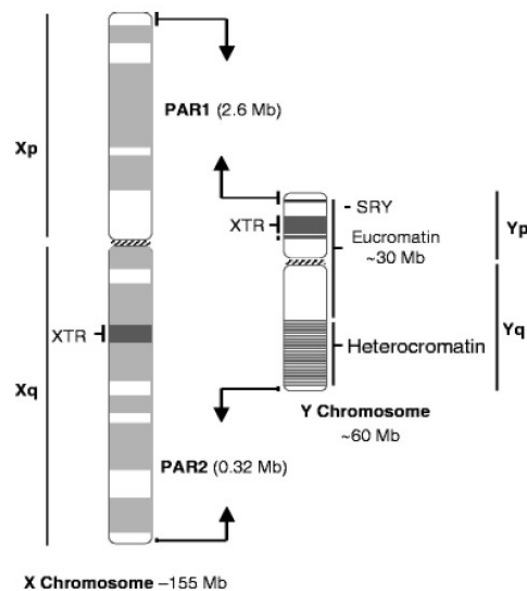


Figura 14: Representação ideográfica das regiões PAR1 e PAR2 no par de cromossomas sexuais. Adaptado de Fernández, 2013.

1.8 Modo de Transmissão do Cromossoma X & Compensação da Dosagem Genética

Hoje em dia, o modo de transmissão do cromossoma X já não é uma incógnita pois desde o início do século XX até à atualidade têm-se desenvolvido diversos e exaustivos estudos, sendo um dos mais importantes triunfos da investigação epigenética (Payer, Lee, & Namekawa, 2011; Wu et al., 2014).

A Epigenética é o ramo da genética que estuda a expressão do DNA, é definida de uma forma simples como sendo a pontuação adicionada ao DNA, ou seja, são as marcas químicas que dão sentido à sequência de pb e que fazem com que as células sejam diferentes entre si, uma vez que todas contêm o mesmo código genético.

Sabe-se que um dos cromossomas X é desativado nas fêmeas por forma a compensar a dosagem de genes entre géneros (Pessia, Engelstädter, & Marais,

2013). Este é um mecanismo drástico uma vez que é um cromossoma inteiro que é silenciado contendo perto de mil genes ativos.

Se não se silenciasse um dos cromossomas X nas fêmeas dos mamíferos placentários estas teriam o dobro de genes em relação aos homens, originando vários distúrbios genéticos.

Existem diversos mecanismos de regulação da dosagem genética do CrX, em função das espécies e em função do desequilíbrio provocado pelos diferentes cariótipos possíveis para a determinação sexual dos organismos.

1.8.1 Lei de Lyon - Inativação do Cromossoma X ou Lyonização.

Em 1949, Barr e Bertram identificaram uma estrutura na periferia do núcleo celular presente apenas nas células somáticas femininas durante a interfase, à qual deram o nome de um dos seus descobridores, o corpo ou corpúsculo de Barr (Barr, M. L., & Bertram, 1949).

Anos mais tarde, em 1959, Susumu Ohno identificou o corpo de Barr, como sendo um CrX inativado, densamente empacotado (Ohno, 1967).

Em 1961, Mary Lyon (Lyon, 1961) colocou pela primeira vez a hipótese da existência de um mecanismo de inativação do cromossoma nos mamíferos placentários o qual mais tarde, depois de demonstrada e exaustivamente fundamentada, passou a designar-se Lei de Lyon, em sua homenagem (Boumil & Lee, 2001).

A Lei de Lyon consiste em dois pressupostos:

- A compensação genética ocorre nas fêmeas através da inativação de um dos cromossomas X, por célula somática diploide, assegurando que apenas um CrX fica ativo;
- Durante a inativação, a escolha do cromossoma X é aleatória.

Por exemplo:

- XX mais 44 autossomas – fêmea diploide normal - um (1) X inativo
- XY mais 44 autossomas – macho diploide normal - nenhum (0) X inativo.
- XXX mais 44 autossomas – indivíduos com Trissomia X – dois (2) X inativos.
- XXY mais 44 autossomas – indivíduos com síndrome Klinefelter – um (1) X inativo.
- XXXX mais 88 autossomas – fêmea com células tetraploides - dois (2) X inativos.

As mulheres têm um cariótipo XX, pois herdam um cromossoma X da mãe, X_m , e um cromossoma X do pai, X_p logo o cariótipo feminino é igual a X_mX_p . E o cariótipo do masculino é XY, nos homens o X é sempre herdado pela via maternal, ou seja, X_m e o Y é sempre herdado pela via paternal, isto é, Y_p , assim sendo, o cariótipo masculino é sempre X_mY_p .

Ao silenciar um dos cromossomas nas mulheres, estas passam a ter dois tipos de células, umas são X_p ativo e outras X_m ativo, com produção de proteínas expressas a partir dos genes do respetivo CrX ativo, assim sendo, as mulheres são geneticamente um mosaico de dois tipos células, por isso têm uma maior variedade genética e, conseqüentemente, de proteínas, ao passo que os homens são compostos exclusivamente pela expressão plena de apenas um cromossoma X, o X_m , sendo todas as células equivalentes pois todas as células têm o mesmo genótipo ativo, X_mY_p .

O X_m sofre recombinação com o outro cromossoma X da mãe durante a meiose ocorrida na formação dos óvulos, pois porta-se como um autossoma; o X_p quase não sofre recombinação durante a espermatogénese, exceto com o cromossoma Y pelas regiões PAR. Assim sendo, os perfis genéticos femininos são homo ou heterozigóticos para cada *loci*, podendo apresentar um ou dois alelos para cada *locus*, respetivamente. E os perfis masculinos são sempre hemizigóticos apresentando apenas um alelo para cada *locus*, formando um perfil haplotípico. Este modo de transmissão confere, ao cromossoma X, uma fonte excecional de informação para estudos de ancestralidade (sobretudo por intermédio dos perfis genéticos masculinos), estudos de diversidade genética populacional (pela sua

faceta autossômica nas mulheres) e estudos forenses para identificação humana e de paternidade, sobretudo em trios mãe-filha-“?” em que não há amostra do pai mas sim da avó paterna.

1.8.2 Mecanismo de Inativação do Cromossoma X

A inativação do CrX é um mecanismo epigenético, cujo objetivo é fazer a compensação de dose, de forma que machos e fêmeas tenham a mesma dose de genes ativos existentes num CrX, por célula diploide.

O ciclo de Inativação do Cromossoma X ocorre em 4 fases (Lee & Bartolomei, 2013, ver Figura 15):

- I. Após a fecundação e formação do ovo, até à formação do blastocisto, o genoma contém informação prévia ou *imprinted* para inativar apenas o CrX paternal, isto é, X_p .
- II. Na formação do blastocisto todas as células voltam a ativar o X_p , ficando os dois CrX novamente ativos, exceto nos tecidos que irão formar a placenta, os quais manterão o X_p inativo ao longo de todo o desenvolvimento intrauterino.
- III. No embrião, a inativação aleatória dá-se a partir da formação do epiblasto e mantém-se nas células somáticas ao longo da vida do mamífero placentário.
- IV. A inativação é novamente removida durante a formação das células germinativas primordiais e durante a embriogénese (quarta semana de gestação), as quais darão origem aos futuros gametas.

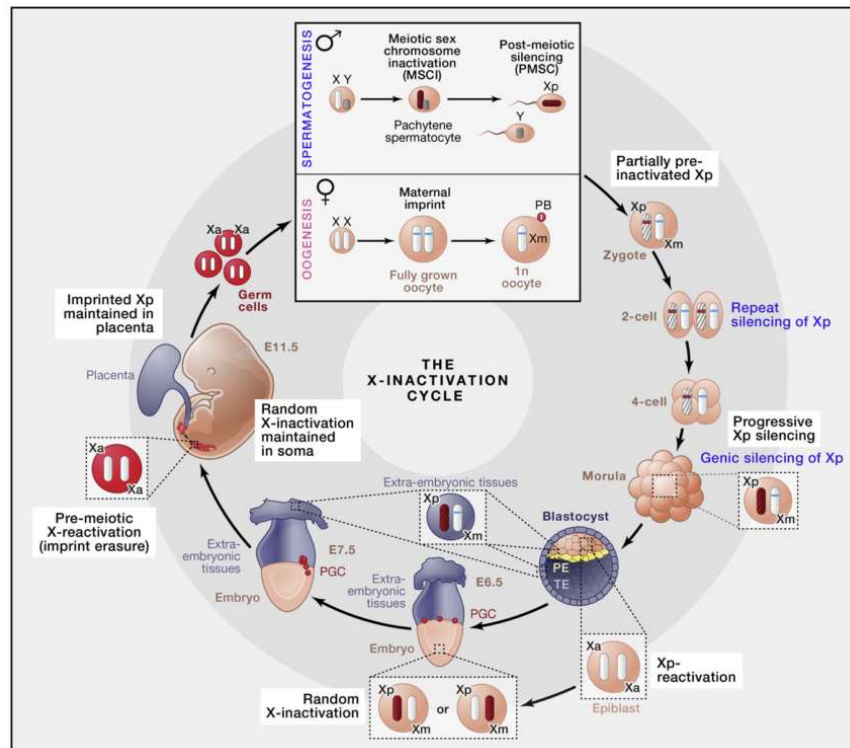


Figura 15: Ciclo de inativação e reativação do CrX.
Adaptado de Lee & Bartolomei, 2013.

A Inativação do Cromossoma X ocorre em 6 etapas (Jeon, Sarma, & Lee, 2012; Nora & Heard, 2009; Payer et al., 2011; Payer & Lee, 2008):

- I. Contagem do número de cromossomas X em relação ao número de autossomas por célula diploide; razão entre X:A vai determinar o grau de inativação necessário.
- II. Escolha, qual dos dois cromossomas X a inativar.
- III. Iniciação da inativação ou silenciamento do cromossoma X escolhido, através da expressão do *Xist*, a partir do X Inactivation Centre (XIC).
- IV. Estender, alastrar ou difundir o silenciamento ao longo de todo o cromossoma em cis.
- V. Estabelecer a inativação, desligando a expressão de *Xist*, após ter coberto praticamente todo o cromossoma.
- VI. Manutenção da Inativação do Cromossoma X, estabilizando-a ao longa da vida da célula e transmitindo-a à sua prol.

1.9 Polimorfismos do Cromossoma X vs Genética Populacional.

Atualmente os polimorfismos mais utilizados em genética populacional para fins forenses ainda são os STRs, tendo caído em desuso polimorfismos obtidos por RFLP, por necessitar da recolha de uma amostra com maior quantidade de DNA. Os STRs permitem a análise de amostras com pouco DNA e/ou degradado, o procedimento é mais barato e eficaz, contudo já começam a perder terreno para os SNP, atualmente com grande destaque na genética forense com aplicação na determinação de fenótipos de indivíduos suspeitos (Ruiz et al., 2013), entre outras aplicações.

Durante as duas últimas décadas, os STR localizados nos autossomas ganharam uma enorme relevância, pois são facilmente estudados por PCR, permitindo genotipar mais de um *locus* em simultâneo, uniformizar procedimentos e realizar comparação de resultados entre laboratórios, utilizar amostras mais pequenas e até decompostas.

Em resumo, os STRs autossómicos quase substituíram o uso de outro tipo de marcadores genéticos para fins forenses (grupos sanguíneos, identificação de proteínas), sendo utilizados na maioria dos casos de identificação humana, testes de paternidade, bem como em casos de análise de grau de parentesco.

Os STRs do CrX (Figura 16) surgiram num contexto complementar ao uso dos STRs autossómicos, devido ao seu modo de transmissão, permitindo obter resultados mais conclusivos e tirar dúvidas em situações para as quais os STRs autossómicos não dão uma resposta irrefutável (Gomes et al., 2012).

Para o efeito foi necessário realizar estudos para localizar os STRs no CrX e sequenciá-los a fim de conhecer a sua composição e ordem de sequências de repetição; conhecer o seu desequilíbrio de ligação (saber se estão localizados em zonas de recombinação), fazer o seu mapeamento, se são independentes entre si ou se são transmitidos associados em bloco, permitindo a construção de haplótipos e, por fim, calcular a razão de verosimilhança a qual permite ter uma ideia com grau de confiança sobre a probabilidade de serem ou não independentes entre si (BGBX, 2014).

Devido ao modo de transmissão, alguns parâmetros estatísticos só se podem calcular para perfis femininos e outros para perfis masculinos (perfis hemizigóticos, pois só têm um CrX).

A conjugação da informação acima referida permite escolher quais os melhores STRs para fins forenses.

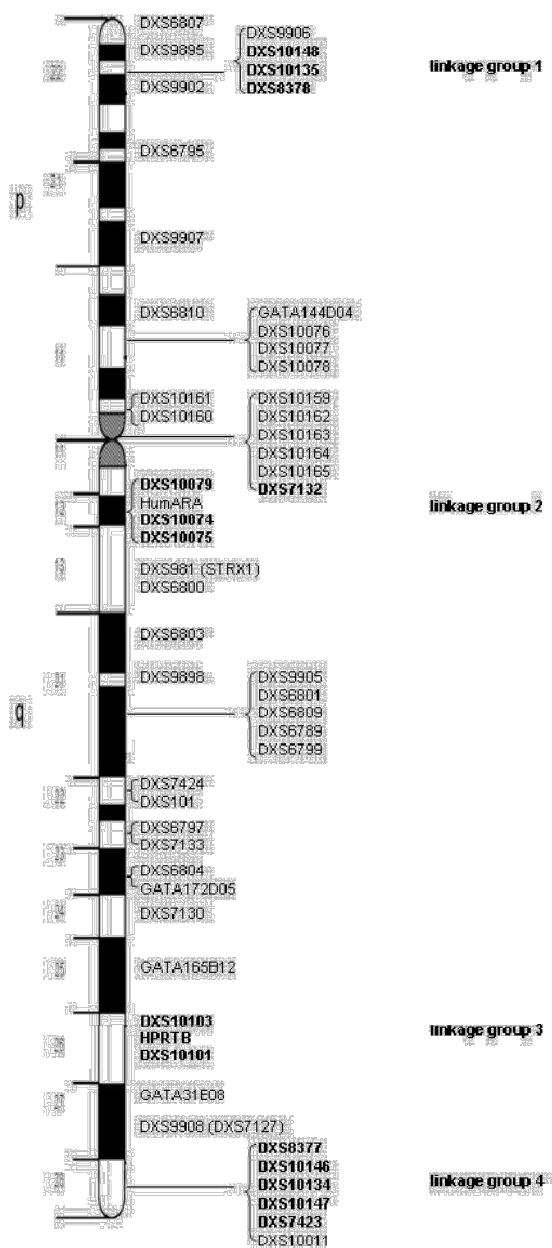


Figura 16: Ideograma do Cromossoma X. Mapa dos 54 loci STR estudados para fins forenses. Imagem adaptada do sítio online do CrX, (ChrX-STR-Org, 2014).

Posta esta primeira etapa, que se conclui com a seleção de STRs, segue-se a segunda etapa que consiste na determinação da sua frequência nas populações, que só é possível obter através dum estudo metódico, trabalho árduo e exaustivo que compreende a recolha de amostras, extração, quantificação, amplificação e genotipagem por eletroforese capilar (por vezes sequenciação, no caso da descoberta de novos alelos).

Uma boa parte dos estudos populacionais realizados pelo mundo inteiro estão disponíveis *online* no site internacional do Cromossoma X, mas de origem e gestão alemã, – “Forensic Chrx Research” (ChrX-STR-Org, 2014), que inclusive disponibiliza ferramentas de cálculo, entre outras informações. E no site brasileiro – “Banco Genético Brasileiro do Cromossomo X” (BGBX, 2014).

1.9.1 Polimorfismos de cromossoma X em contexto forense.

Os STRs do CrX surgiram num contexto complementar ao uso dos STRs autossómicos com o objetivo de:

- Identificar indivíduos;
- Determinar a paternidade e outros graus de parentesco:
 - Envolvendo indivíduos aparentados – em diversos contextos, por exemplo: violação da filha pelo pai ou outro parente;
- Testes de maternidade – embriões ou fetos e restos placentários abandonados;
- Testes indiretos de paternidade – quando não existe amostra disponível do putativo pai.

Os parâmetros que permitem decidir se determinados perfis permitem identificar tanto o indivíduo como a sua linhagem são obtidos a partir das frequências em que os seus alelos se dispersam na população, logo não há Genética Forense sem Genética Populacional.

1.10 Estudos populacionais na região do sudeste asiático

A caracterização da evolução genética ao longo do tempo e no espaço geográfico do sudeste asiático ainda é reduzida, mas tem vindo a ser alvo de diversos estudos

(Figura 17) à medida que vão surgindo novos métodos, equipamentos com capacidade de processar mais amostras e agilizar os procedimentos laboratoriais, assim como, à medida que se vão desenvolvendo novas ferramentas para a análise de dados.

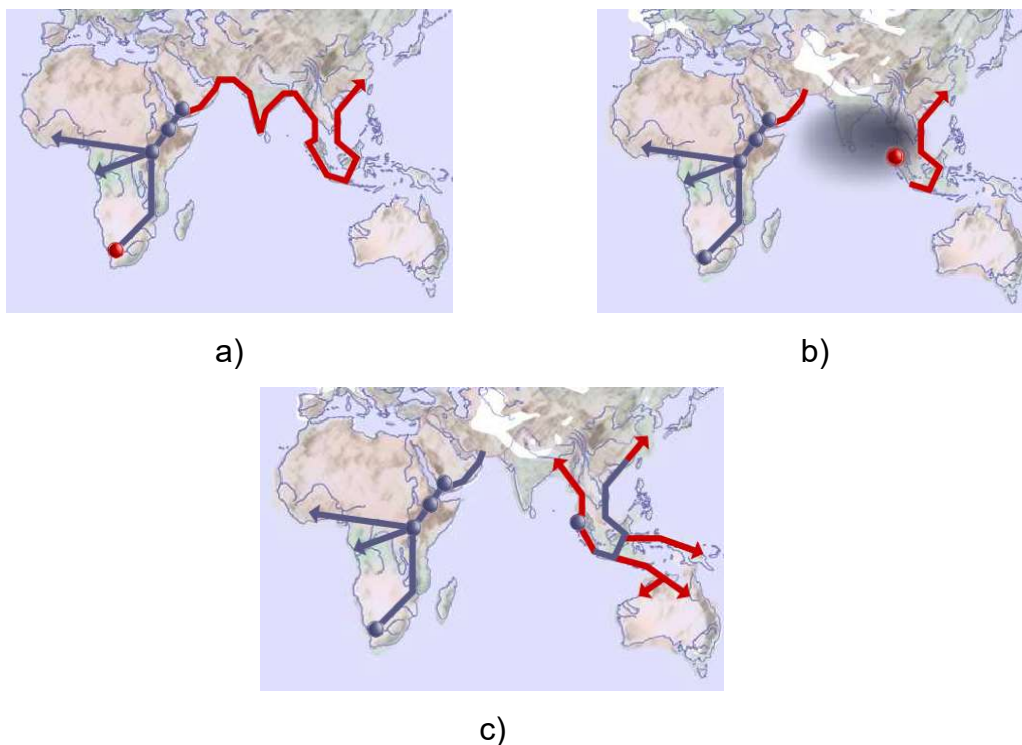


Figura 17: Mapa das primeiras migrações humanas em direção ao sudeste asiático.

Gráficos foram desenhados a partir da conjugação de estudos do DNA mitocondrial e do cromossoma Y.

- a) Representa um dos dois eventos da saída de África pelos primeiros *Homo sapiens*, representa o caminho feito à 85-75 kya (kya= mil anos), tendo como origem o sudeste de África.
- b) Representa um evento catastrófico, ocorrido à 74 kya, o qual limitou os movimentos migratórios, trata-se da erupção do Monte Toba, em Sumatra, provocando um inverno nuclear na região tendo reduzido a população daquela área para perto de 10 000 habitantes.
- c) Com o decorrer do tempo, na eminência da luta pela sobrevivência, seguiu-se a expansão para o nordeste e sudeste daquela região do globo terrestre. A imagem c) descreve o caminho percorrido entre 74-64 kya levando ao início do povoamento de Timor e Austrália.

Adaptado de Oppenheimer, 2014.

Recentemente, foi publicado um artigo atualizado sobre a diversidade genética na região de Sunda (constituída pela península Malaia e parte do Sudeste do continente asiático, as grandes ilhas de Bornéu, Java e Sumatra sendo delimitada a sul pela linha de Wallace) e Sahul (também denominada por Meganésia, região geográfica correspondente à Austrália, Nova Guiné, Tasmânia, Seram e ilhas adjacentes, por exemplo: a ilha de Timor) com foco detalhado na população de

Timor-Leste é relativo à época geológica do Pleistoceno, que antecede a atual época em que vivemos, o Holoceno, período Quaternário, da era Cenozóica, do eón Fanerozoico (Gomes et al., 2015). Este estudo descreve a colonização humana daquela região geográfica com base no estudo do DNA mitocondrial (Figura 18).

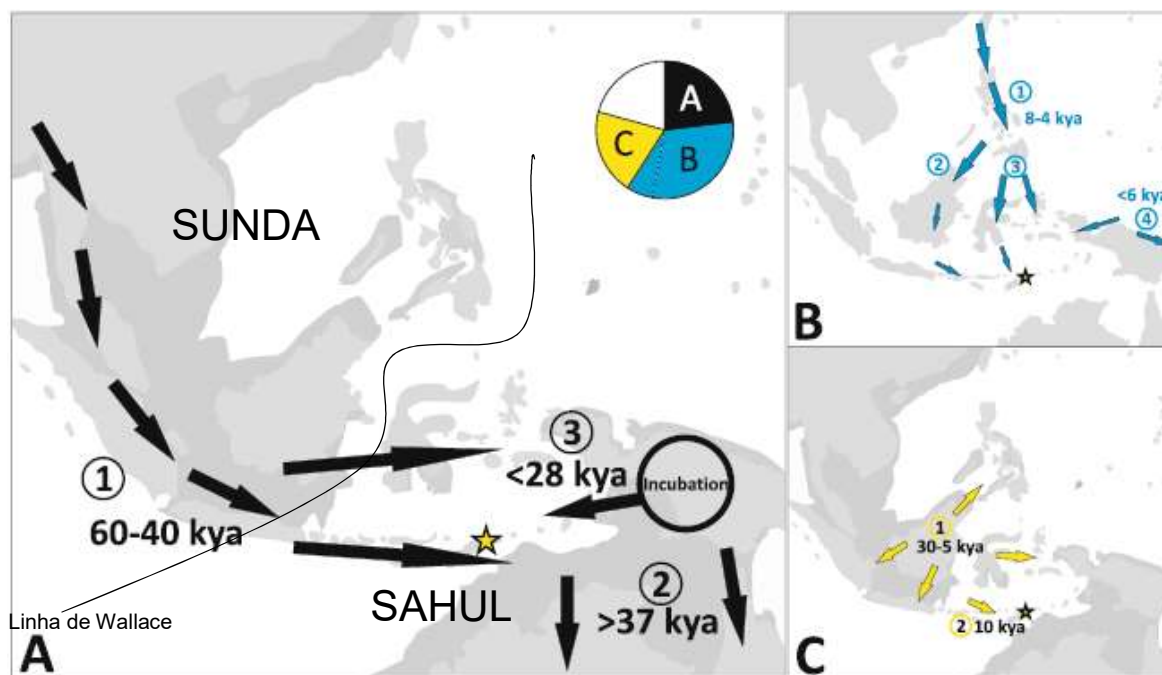


Figura 18: Movimentos migratórios que contribuíram para a pool genética de Timor-Leste.
A-primeiros movimentos migratórios em direção à região marítima do sudeste asiático, ocorridos entre 60-28 kya.
B-Movimentos entre 8-4 kya, Holoceno.
C-Movimentos entre 30-5 kya, com destaque para a colonização da Indonésia há cerca de 10 kya, assinalada com o número 2.

Adaptado de S. M. Gomes et al., 2015.

Atualmente, já existem estudos de diversidade genética obtidos com STRs do CrX naquela zona do globo, relativos apenas a populações chinesas e japonesas, com amostras masculinas, obviamente, pois apenas pela via paterna se podem estabelecer as linhagens, por conterem apenas um CrX permitindo a criação de haplótipos (ChrX-STR-Org, 2014).

Existem também estudos antropológicos (Roque, 2008) e linguísticos (Center for Regional Studies, 1997; Soares, 2014) algo consistentes com os estudos genéticos, mas pouco exatos. Pouca mais informação existe para descrever com rigor a origem e evolução daquelas populações complexas em resultado da sua natureza geográfica. O facto de ser uma ilha contribui para um certo grau de isolamento, pois

o mar constitui uma barreira natural, mas é uma ilha rodeada por várias outras ilhas, servindo como ponto de passagem. Os seus recursos naturais, nomeadamente, a madeira de sândalo e atualmente o petróleo, entre outros, atraíram de povos, originando uma grande diversidade genética com características muito específicas em consequência das diversas invasões e movimentos migratórios.

Todos os restantes estudos genéticos de ancestralidade são mais generalistas e menos evidentes, e correspondem às mesmas épocas geológicas Pleistoceno e Holoceno, baseados em escassas amostras de partes de esqueletos, desde um pré-molar a um pequeno osso de um dedo, assim como, diversos achados arqueológicos. Contudo estes evidenciam a possibilidade dos primeiros *Homo sapiens* se terem cruzado com *Homo erectus* e neandertais, na região junto à Índia com denisovanos, e mais a sul, já junto da linha de Wallace, com o *Homo floresiensis*, contudo ainda não há garantia que este tenha sido um hominídeo, visto existirem evidências se teria sido um *Homo sapiens* com problemas anatómicos e fisiológicos. Como nota, diversos estudos genéticos permitem concluir que a diferença entre os dados obtidos do esqueleto mais recente de *Homo erectus* e o de um *Homo sapiens* mais antigo é de cerca de 100 kya (Dennell & Petraglia, 2012).

Capítulo III Objetivos

2 Enquadramento

A população de Timor-Leste é constituída por 1 066 582 pessoas sendo 541 147 homens e 525 435 mulheres, de acordo com os resultados do Censos 2010, que abrangeu 95% da população (Timor-Leste, 2010).

Foi realizado um novo Censo e terceiro realizado no país, entre 11 e 25 de Julho de 2015, cujos resultados estão a ser apurados.

Atualmente, a população timorense está caracterizada por diversos marcadores genéticos do tipo STRs autossómicos com o AmpF ℓ STR Identifiler (L Souto et al., 2004, 2005) e com o PowerPlex Fusion System $\text{\textcircled{R}}$ (L. Souto, Tavares, Moreira, & Corte-Real, 2013), com STRs do CrY com o PowerPlex $\text{\textcircled{R}}$ Y System kit (Souto, Gusmão, et al., 2006; Souto et al., 2006; Souto, Amorim et al., 2006), marcadores do DNA mitocondrial (HVI *control region* (16050-16400) amplificada com os *primers* L15997/H16401) (L. Souto, Rocha, et al., 2006), SNPs com o SNPforID 52-plex *panel* (Santos et al., 2011) e STRs do CrX com o *Investigator* $\text{\textcircled{R}}$ Argus X-12 ((Moreira, Costa, Tavares, & Souto, 2015)

Tabela 1: Estudos genéticos disponíveis sobre amostras da população de Timor-Leste.

Ano	Amostra T:M/F	STR	Multiplex	Literatura
2014	149 : 101/48 Saliva/FTA $\text{\textcircled{R}}$, cartões	12 X-STRs	Investigator Argus-X12	(Moreira et al., 2015)
2013	100 Saliva/FTA $\text{\textcircled{R}}$, cartões	24 STRs AS	PowerPlex FusionSystem $\text{\textcircled{R}}$ (CODIS+ESS core <i>loci</i>)	(Souto et al., 2013)
2011	46	52 SNPs AS	SNPforID 52-plex panel	(Santos et al., 2011)
2006	(i) 138 (M) Sangue+saliva/FTA $\text{\textcircled{R}}$ cartões; (ii) 346 (M) Sangue/FTA $\text{\textcircled{R}}$, cartões; (iii) 342 (M) Sangue/FTA $\text{\textcircled{R}}$, cartões	12 STRs do CrY [(i), (ii) e (iii)]	Power-Plex $\text{\textcircled{R}}$ Y System [(i), (ii) e (iii)]	(i) (Souto et al., 2006); (ii) (Souto, Gusmão, et al., 2006); (iii) (Souto et al., 2006).
2006	133 Saliva+Sangue/FTA $\text{\textcircled{R}}$, cartões	Região controlo HVI (16050-16400)	Primers L15997/H16401)	(Souto et al., 2006)
2005	186 inclui os 107 abaixo reportados Saliva+Sangue/FTA $\text{\textcircled{R}}$, cartões	15 STRs AS	AmpF ℓ STR Identifiler	(Souto et al., 2005)
2004	107 Saliva+Sangue/FTA $\text{\textcircled{R}}$, cartões	15 STRs AS	AmpF ℓ STR Identifiler TM	(Souto et al., 2004)

Em genética, multiplex é a designação para um sistema de *primers* que permite genotipar mais do que um loci em simultâneo.

2.1 Objetivo Geral

Caracterização genética de uma amostra de 40 indivíduos, 35 homens e 5 mulheres, da população de Timor-Leste, com o X-STR Decaplex da GHEP-ISFG (“GHEP-ISFG,” 1995), com referência aos perfis genéticos 9947A (DNA humano feminino) e 9948 (DNA humano masculino).

Comparação do sistema X-STR Decaplex com o sistema Investigator Argus X-12.

2.1.1 Objetivo Específico

Genotipagem com o sistema decaplex desenvolvido pelo GHEP-ISFG para os seguintes *loci* do cromossoma X: DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789. Cálculo dos parâmetros forenses.

Comparação dos resultados obtidos com resultados de uma amostra equivalente, genotipada com o sistema *Investigator*® Argus X-12. Este utiliza 24 *primers* para os seguintes *loci* do cromossoma X: DXS10103, DXS8378, DXS7132, DXS10134, DXS10074, DXS10101, DXS10135, DXS7423, DXS10146, DXS10079, HPRTB e DXS10148.

Capítulo IV Material & Métodos.

3 Amostragem.

As amostras utilizadas pertencem à coleção de Timor-Leste do Laboratório de Genética Aplicada do Departamento de Biologia da Universidade de Aveiro. A origem e grupo etnolinguístico dos indivíduos intervenientes, bem como dos seus pais e avós, foram obtidos através da realização de inquéritos, durante a recolha das amostras. Estas foram recolhidas entre voluntários de acordo com critérios básicos de consentimento informado e escrito.

As amostras de sangue e saliva foram colhidas para cartões de FTA® e o DNA extraído pelo método de Chelex® (Walsh et al., 1991).

O ensaio com o X-STR Decaplex do GHEP-IFGS foi efetuado sobre 40 indivíduos, dos quais 35 homens e 5 mulheres. As amostras de sangue e saliva foram recolhidas em cartões FTA®.

O GHEP-IFGS é uma associação cívica de carácter científico, sem fins lucrativos, integrada na *International Forensic Genetic Science*, com sede em Madrid, que desenvolve atividades em Portugal, Espanha e na América Latina, apesar de poder desenvolver atividades noutras áreas geográficas. O GHEP-ISFG tem por objetivo principal promover e contribuir para o desenvolvimento e disseminação do conhecimento científico na área da genética forense e áreas relacionadas. O GHEP-ISFG promove relações de trabalho e colabora com grupos científicos de diversas universidades, os primers do X-STR Decaplex foram cedidos mediante um anterior programa de colaboração com o grupo do Professor Luis Souto.

Em simultâneo genotipou-se com o Investigator Argus X-12 uma amostra composta por 134 indivíduos, dos quais 86 homens e 48 mulheres. Esta amostra de DNA provém de saliva recolhida por zaragatoas e depositada em cartões FTA®.

O Investigator Argus X-12 é o nome comercial de um kit que contém os *primers* para 12 loci do CrX, desenvolvido pela Qiagen.

3.1 Extração e Quantificação de DNA genómico

O DNA foi extraído dos cartões de FTA® com um micro *punch*, retirando pedaços circulares de no máximo 9 mm² e colocados em *eppendorfs* de 1,5 ml (tantos

ependorfs quantas as amostras necessárias), de seguida aplicou-se o método de Chelex (Walsh et al., 1991), de acordo com o seguinte procedimento:

para cada *ependorf* pipetar 1 ml de H₂O ultrapura (Tipo I) ou Mili-Q®;

agitar e incubar à temperatura ambiente 15-30 minutos;

centrifugar a 14000rpm durante 2-3 minutos e retirar o sobrenadante, deixando aproximadamente 25 µl;

adicionar 175 µl de solução de Chelex100 a 5% e vortexar;

incubar 15-30 minutos a 56 °C e misturar no vórtex 5-10 segundos;

Furar as tampas dos tubos e incubar num banho em ebulição (90-99°C) durante 8 minutos (para abrir as células);

Vortexar 5-10 segundos;

Centrifugar 2-3 minutos a 14000rpm (para separar a resina da solução de DNA);

Usar diretamente em PCR, após validação por quantificação.

A PCR ou reação química em cadeia da polimerase (Taq-polimerase) permite fazer várias cópia inteiras ou de partes de DNA (amplificar), este processo de amplificação foi descrito pela primeira vez por Kary Mullis (Mullis & Faloona, 1987), em 1983, e valeu-lhe o Prémio Nobel da Química em 1993.

A quantificação permite verificar se a extração foi eficaz e se o DNA está disponível na quantidade necessária para amplificar com os sistemas multiplex para STRs habitualmente comercializados.

A quantidade de DNA deve estar entre 0,5 ng a 2,0 ng, uma vez que pouco DNA faz com que haja perda de alelos e demasiado DNA permite a formação de stutters e picos fora da escala eletroforética (Butler, 2012).

Assim, se for necessário, procede-se a nova extração ou à diluição das amostras de DNA para, finalmente, se proceder à amplificação.

O equipamento utilizado para a quantificação foi Qubit® Fluorometer da Invitrogen™ recorrendo ao Quant-iT™ dsDNA BR Assay Kit, procedendo de acordo com as instruções do fabricante, da seguinte forma:

Preparar as soluções do kit Quant-iT™ dsDNA BR da Assay, de acordo com as instruções do fabricante, serão preparados tubos de 0,5ml, dois para os standards e um para a amostra;

Agitar as amostras e os padrões durante 2-3 minutos;

Incubar, durante dois minutos, à temperatura ambiente;

Calibrar o Qubit™ fluorometer com as soluções standard de DNA (o kit Quant-iT™ dsDNA BR contém 2 standards);

Inserir a amostra no Qubit fluorometer e quantificar.

As amostras, uma vez extraídas e/ou quantificadas, devem ser armazenadas a, aproximadamente, -20 °C até à altura da sua utilização para amplificação.

3.2 Amplificação X-STR Decaplex.

A amplificação do DNA extraído foi levada a cabo utilizando o multiplex constituído pelo sistema de *primers* descrito na Tabela 2 **Erro! A origem da referência não foi encontrada.**, para os *loci* representados na Figura 19, com o Qiagen® Multiplex PCR kit (Technologies, 2010), o qual contém a solução comercial Master Mix (composta por HotStartTaq DNA Polymerase, MgCl₂, dNTPs e solução tampão para PCR) e água esterilizada, também designada por água PCR-grade ou RNase free.

Tabela 2: Tabela de Primers para o X-STR Decaplex GHEP-ISFG, Leonor Gusmão et al., 2009

Locus	Marcador de Fluorescência	Sequência dos primers (5'→3')	T emp (°C)	Tamanho (bp)	Localização citogenética	Localização Física (Mb) NCBI 36
DXS8378	6-FAM	F:TTAGGCAACCCGGTGGTCC R:ACAAGAACGAACTCAACTC	60°C	109-141	Xp22.31	9,330
DXS9898		F:CGAGCACACCTACAAAAGCTG R:TAGGCTCACCTCACTGAGCA	60°C	150-180	Xq21.31	87,682
DXS7133		F:CACTTCCAAAAGGGGAAAAA R:ACTTGACTTGGTGGGAGGAA	56°C	204-265	Xq22.3	108,928
GATA31E08		F:GCAAGGGGAGAAGGCTAGAA R:TCAGCTGACAGAGCAGAGAGA	63°C	267-315	Xq27.1	140,062
GATA172D05	VIC	F:TAGTGGTGATGGTTGCACAG R:ATAATTGAAAGCCCGGATTC	58°C	113-155	Xq23	113,061
DXS7423		F:GTCTTCCTGTCATCTCCCAAC R:TAGCTTAGCCTGGCACATA	62°C	173-204	Xq28	149,460
DXS6809		F:TCCATCTTTCTCTGAACCTTCC R:TGCTTAGGCTGATGTGAGG	58°C	229-280	Xq21.33	94,4825
DXS7132		F:TCCCCTCTCATCTATCTGACTG R:CACTCCTGGTGCCAACTCT	56°C	130-162	Xq11.2; centrómero	64,572
DXS9902	NED	F:CTGGGTGAAGAGAAGCAGGA R:GGCAATACACATTATATCAGGA	60°C	130-162.8	Xp22.2	15,234
DXS6789		F:CTTCATTATGTGCTGGGGTAAA F:ACCTCGTGATCATGTAAGTTGG	60°C	249-297	Xq21.33	95,336

Nota: Cores dos painéis, 6-FAM, azul; VIC, verde e NED, amarelo. F e R, *forward primer* e *reverse primer*, respetivamente.

O procedimento é o indicado pelo GHEP-ISFG (Gusmão et al., 2008; Gusmão et al., 2009), preparando a solução mistura, conforme na Tabela 3, e a mistura de amplificação, de acordo com a proporção:

9,5 µL de solução mistura (ver Tabela 3), aos quais adicionar

0,5 µL de amostra DNA previamente extraída ('DNA template').

Preparar um volume de solução mistura igual ao número de amostras a genotipar por ensaio, tal que, se o número de amostra for 40, o volume total de solução mistura, deverá ser suficiente para um pouco mais de 40, de forma a haver um ligeiro excesso de solução mistura, por exemplo, 9,5 µL x 50 = 475 µL, o excedente permite proceder à preparação do controlo negativo, das amostras de DNA padrão necessárias e reserva de solução para o caso de ser necessário fazer repetições durante o ensaio.

Preparar nova solução mistura, preparar novo controlo negativo e genotipar amostra padrão 9947A (Promega, USA) e 9948 (Promega, USA), para cada ensaio.

O controlo negativo é preparado substituindo o DNA template por água PCR-grade.

Tabela 3: Preparação das misturas reacionais.

	Composição	Volume (μL)	
Solução Mistura	Master Mix kit Qiagen	5	9,5
	PimerMix GHEP-ISFG	1	
	H ₂ O PCR-grade kit Qiagen	3,5	
	DNA _{template}	0,5	

A solução de mistura de *primers*, designada na Tabela 3, como *PrimerMix*, inclui todos os *primers* descritos na **Erro! A origem da referência não foi encontrada.**, numa concentração de 2 μM.

Proceder do seguinte modo para preparar a ‘Primer Mix’:

preparar soluções de 100 μM de cada *primer*

adicionar uma alíquota de 5 μL de cada solução de *primer* (5x20 *primers*) e diluir em 150 μL de água esterilizada ou PCR-grade, perfazendo 250 μL de volume total.

Após a preparação das amostras proceder de acordo com o ciclo de amplificação térmica. Para o efeito utilizar um termociclador, neste caso o iCycler™ da BioRad, de acordo com o programa descrito na tabela que se segue, Tabela 4.

Tabela 4: Ciclo de amplificação para PCR.

	Temperatura	tempo	n repetições
HotStart (desnaturação inicial)	95°C	15min	1x
Desnaturação	94°C	30 seg	10x
Hibridação	60°C	90 seg	
Extensão	72°C	60 seg	
Desnaturação	94°C	30 seg	20x
Hibridação	58°C	90 seg	
Extensão	72°C	60 seg	
Extensão final	60°C	60 min	1x
Repouso	4°C		∞

A localização dos *loci* X-STR do Decaplex do GHEP-ISFG está representada na seguinte figura:

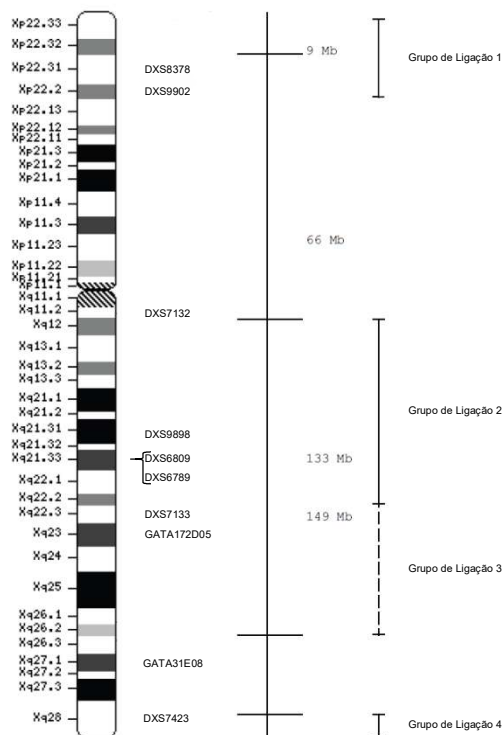


Figura 19: Marcadores do X-STR Decaplex do GHEP-ISFG. Imagem adaptada do sítio online do NCBI, (NCBI, 2015).

O sistema de *primers* que flanqueiam os diversos *loci* foram estudados, otimizados e validados pelo GHEP-ISFG, estão caracterizados conforme o descrito na literatura (Leonor Gusmão et al., 2009), e, fazem parte da composição da solução mistura de *primers*, designada por *PrimerMix* na Tabela 3.

3.3 Amplificação *Investigator*® Argus X-12.

A amplificação por PCR com o sistema Argus X-12, kit *Investigator*® Argus X-12 da Qiagen, foi efetuada de acordo com as instruções descritas no manual do fabricante (Qiagen, 2010).

A localização dos *loci* X-STR está de acordo com a Figura 20.

A Figura 20 representa o mapa dos 12 X-STRs genotipados pelo *Investigator*® Argus X-12.

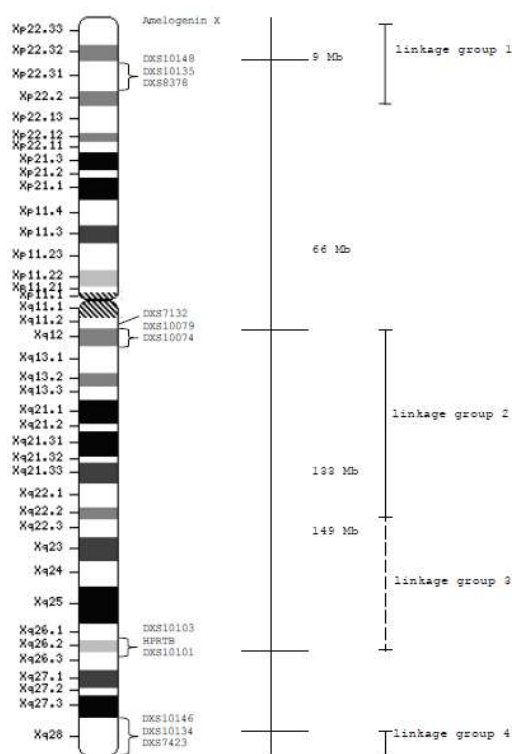


Figura 20: Marcadores do *Investigator*® Argus X-12.

Imagem adaptada do Manual de Instruções, *Investigator Argus X-12 Handbook* da Qiagen, versão de 2010.

O sistema de *primers* está caracterizado no manual do kit Argus X-12 (Qiagen, 2010).

3.4 Eletroforese em Gel de Agarose - AGE.

As amostras amplificadas foram previamente examinadas por AGE (Figura 21), a fim de confirmar o sucesso da amplificação por PCR, de acordo com o seguinte procedimento:

Preparação do gel de agarose a 2% com corante GelRed™;

Preparação das amostras de acordo com o seguinte procedimento e ordem:

	Marcador de pesos moleculares ('Ladder')	Controlo negativo (C-) & Produto Amplificado
	4 µL água PCR-grade	5 µL de controlo negativo ou produto amplificado
	1 µL de 50 pb DNA Ladder	
	1 µL DNA Loading Dye (6x) azul	1 µL de Orange Dye (6xOrange DNA Loading Dye)
Volume Total	6 µL	6 µL

Condições da corrida: DDP 130 volt; corrente a 400mA, t 25 min, em solução tampão de TBE 1x, igual à utilizada para preparar o gel de agarose.

A eletroforese em gel de agarose é uma técnica de separação de moléculas baseada na interação entre carga e tamanho, onde o gel funciona como um peneiro molecular atrasando as moléculas maiores ao mesmo tempo que o campo elétrico aplicado ao sistema atrai as moléculas mais pequenas para o polo positivo, uma vez que o DNA tem carga negativa (causada pela composição conter em fosfato).

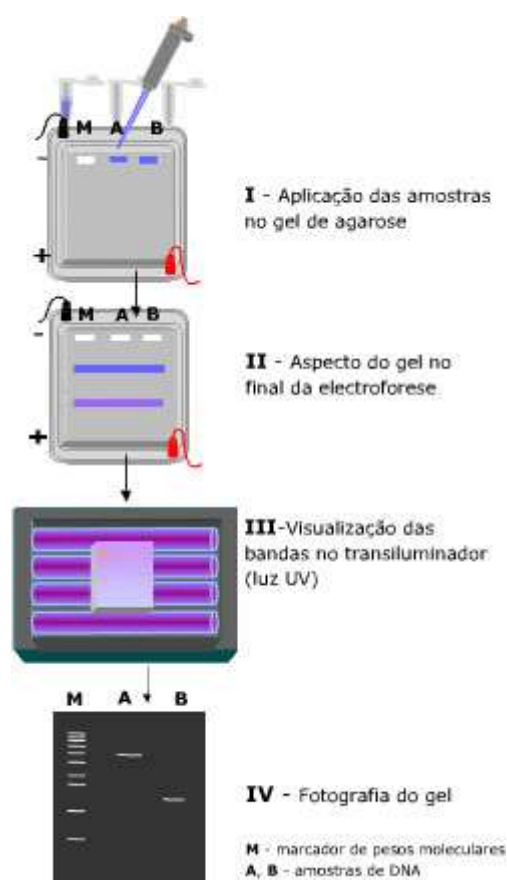


Figura 21: Representação do procedimento de separação do produto amplificado por eletroforese em gel de agarose, AGE.

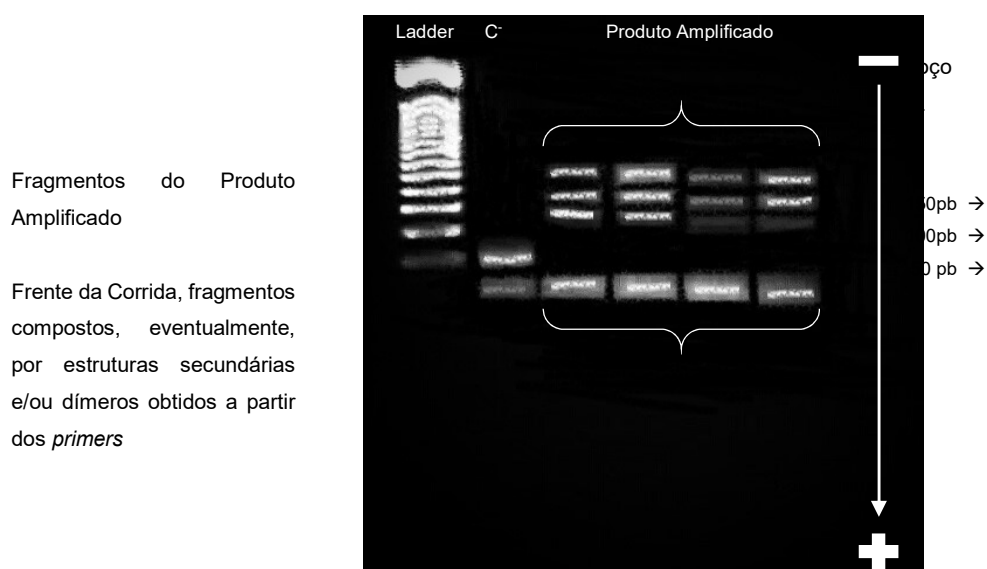


Figura 22: Exemplo de uma eletroforese em gel de agarose obtido após genotipagem com o X-STR do Decaplex GHEP-ISFG.

Resultado observado no transluminador UV, decorridos 25 minutos.

Após confirmação do sucesso da amplificação por PCR (Figura 22), guardam-se as amostras no frigorífico a -20°C até posterior genotipagem por eletroforese capilar.

3.5 Eletroforese Capilar e Genotipagem

A eletroforese capilar baseia-se no mesmo princípio da eletroforese em gel normal, porém o gel e a amostra são transportados dentro de um capilar e os fragmentos são identificados por um detetor (Figura 23).

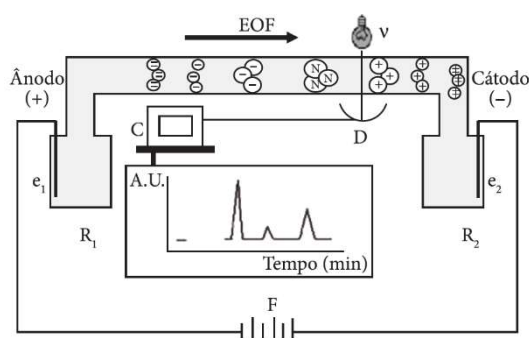


Figura 23: Esquema ilustrativo de um equipamento de Eletroforese capilar.

Onde R1 e R2, e1 e e2 são os reservatórios e eletrodos, respectivamente. F representa a fonte de alta tensão, D é o detetor, C é o computador para obtenção dos dados e EOF representa o fluxo eletrosmótico. Ao centro observa-se um eletroferograma onde se observa a separação dos fragmentos. Cada pico corresponde a um fragmento amplificado. Imagem adaptada de Spudeit, Dolzan, & Micke, 2012.

O produto amplificado ou amplicons, obtido com X-STR Decaplex e com Argus X-12, foram separados por electroforese capilar num sequenciador automático ABI PRISM 3500 Genetic Analyzer® da Applied Biosystems®; a análise de dados foi realizado com o GeneMapper® ID-X 1.2 Software. A eletroforese capilar foi aplicada em condições desnaturantes, utilizando polímero POP4 (Applied Biosystems). Como nota, a eletroforese capilar e genotipagem com o Argus X-12 foram efetuadas de acordo com as condições do manual do fabricante para o kit *Investigator*® Argus X-12 da Qiagen.

As amostras a genotipar foram preparadas da seguinte forma:

10 μL de formamida desionizada+ 1 μL de GeneScan 500LIZ + 1 μL produto amplificado por PCR, em tubos de 0,5 mL com septo próprio.

Misturar no vortex e centrifugar por 15s a 14 000 rpm;

Levar ao termomixer/termobloco a 95°C por 5 min;

Colocar no suporte e levar ao congelador por 5 min, ou deixar no frigorífico a 4°C, até serem analisadas;

Antes de genotipar:

Misturar no vortex e centrifugar por 15s a 14 000 rpm;

Aquecer até atingir a temperatura ambiente.

A análise foi realizada num ABI PRISM 3500 Genetic Analyzer (Applied Biosystems®), utilizando o módulo de corrida “GS STR POP-4 (1ml) G5” e a matriz standard “Matriz Standard Set DS-33” (Applied Biosystems®); esta matriz é formada pelos corantes 6-FAM™ (azul), VIC® (verde), NED™ (amarelo), PET® (vermelho) e LIZ® (cor-de-laranja).

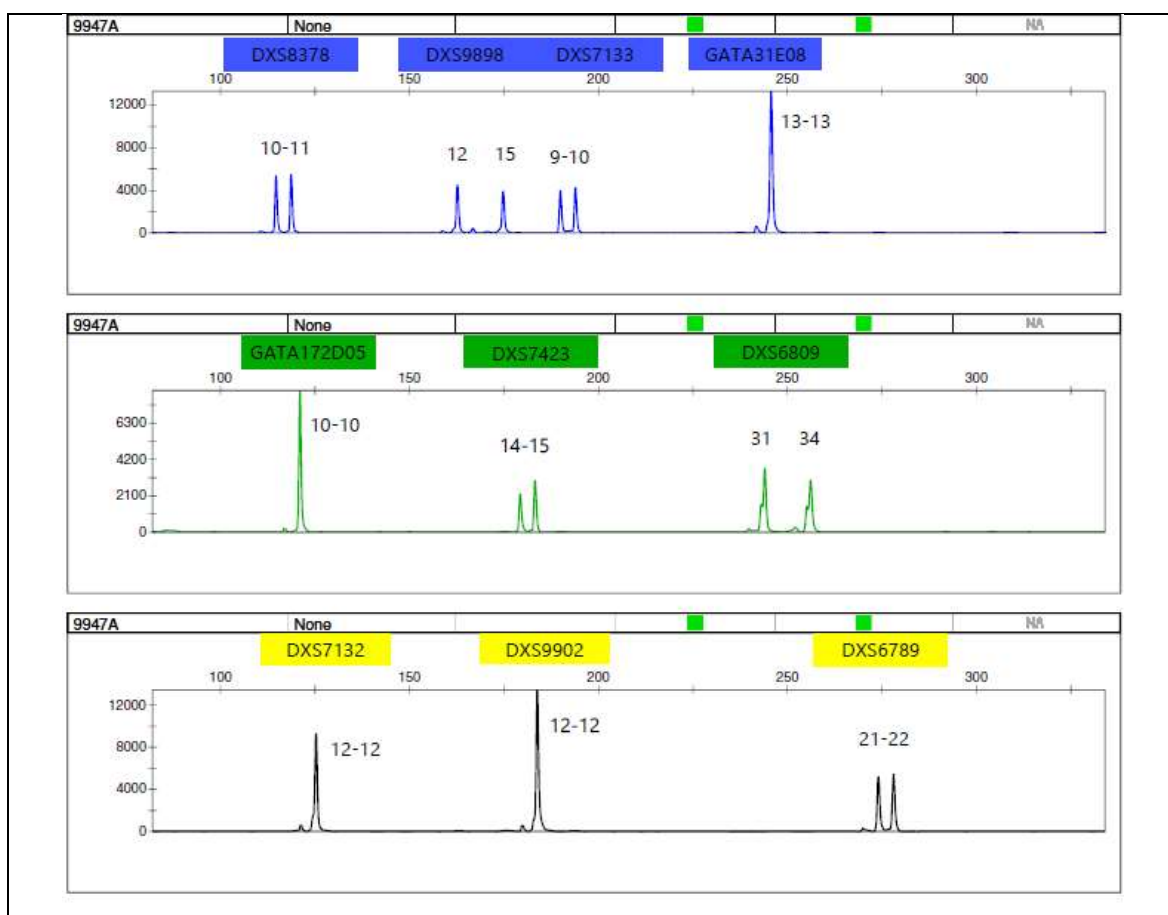


Figura 24: Eletroferograma do perfil genético para o padrão 9947A com o painel X-STR Decaplex do GHEP-ISFG.

A amostra de referência da Figura 24 foi genotipada e os alelos atribuídos de acordo com os resultados obtidos a partir do artigo de referência Leonor Gusmão et al., 2009.

Os resultados foram analisados por comparação com os tamanhos dos fragmentos obtidos para o DNA padrão 9947A (DNA feminino) (ver Figura 24), e 9948 (DNA masculino), conhecendo quais os alelos correspondentes, assim como, por intermédio de uma folha de cálculo Excel, tendo em conta a unidade base de repetição da sequência de pb para cada STR (ver Tabela 5), baseada em estudos prévios devidamente publicados e referidos na literatura.

O tamanho dos fragmentos pode variar de acordo com a qualidade das soluções de *primers*, o rendimento da PCR durante a amplificação, assim como com as flutuações de corrente elétrica do sistema de eletroforese ou sequenciador automático. O uso do padrão interno cujos tamanhos são conhecidos permite reduzir o erro de interpretação durante a classificação manual dos alelos, servindo como termo de comparação. O padrão interno também pode ser designado pelo termo *ladder* por funcionar como uma régua, contendo tamanhos bem definidos de fragmentos não alélicos. O *ladder* alélico é uma solução composta por todos os alelos conhecidos para um dado loci, no caso dos multiplexs contém todos os alelos conhecido para o conjunto de *loci*.

O padrão interno utilizado foi o GS-500LIZ Size Standard, com fragmentos em número e tamanhos conhecidos, conforme já foi referido. Pode haver variação no tamanho detetado de acordo com o padrão interno, por exemplo para o GS-600LIZ o tamanho médio de um fragmento, logo deve utilizar-se sempre o mesmo padrão interno com todas as amostras.

Os resultados da aplicação do kit Argus foram obtidos por comparação com um *ladder* alélico, o DNA padrão XX28, através do software de análise de dados 'GeneScan®' ou 'GeneMapper™ ID'. A análise dos dados obtidos por eletroforese capilar foi realizada com o GeneMapper™ ID-X 1.2 software, que faz parte do sequenciador automático ABI PRISM 3500 Genetic Analyzer (Applied Biosystems®), o qual permite a atribuição automática dos alelos.

Os dados obtidos com o sistema Decaplex foram, posteriormente, cruzados com os dados obtidos com o *Investigator®* Argus X-12 para os três *loci* que ambos os sistemas têm em comum (DXS 8378, DXS7132, DXS7423).

Capítulo V Resultados.

4 Análise dos marcadores X-STRs

Após classificação alélica manual (X-STR Decaplex), locus a locus, tendo em conta as possíveis sequências para cada STR, descritas na Tabela 5, conforme referido anteriormente, ou por classificação digital (Argus X-12), através do software informático do sequenciador automático, isto é, uma vez recolhidos os dados, efetuaram-se diversos cálculos, nomeadamente, parâmetros de estatística descritiva e de inferência estatística, a fim de caracterizar a amostra de perfis genéticos e comparar alguns desses parâmetros com os obtidos para outras populações.

Tabela 5. Caracterização dos fragmentos de amplificação de acordo com o Sistema de primers do X-STR Decaplex do GHEP-ISFG

Locus	Alelo	Fragmento (tamanho)	Sequencia/Tamanho da unidade de repetição	Perfil 9947A	Perfil 9948	Ref
DXS8378	8-14	110-134	PF-N ₁₈ -(CTAT) ₈₋₁₄ -N ₂₀ -PR / 4	10/11	11	(Edelmann et al., 2002)
DXS9898	7-16, 8.3	212-248	PF-N ₁₆ -(TATC) ₂ -ATC-(TATC) ₁₋₅ -(ATC) _{0,1-} (TATC) ₅₋₁₃ -(TATC) ₂ atc(TATC) ₁₋₅ -(ATC) _{0,1-} (TATC) ₅₋₁₃ -N ₉₅ -PR / 4 ou 4+0.1(ATC)	12/15	13	(Gomes et al., 2007)
DXS7133	9-14	204-265 ou 180-200	PF-(ATAG) ₉₋₁₄ -N ₃₆ -PR ou (ATAG) _n / 4	9/10	11	(Gusmão et al., 2009)
GATA31E08	5-15	267-315 ou 240-260	PF-(AGGG) _{0,2-3} -(AGAT) _{5,7-13} -PR / 4	13/13	12	(Gusmão et al., 2009)
GATA172D05	5-12	108-136 ou 113-155	PF-N ₅ -(TAGA) ₅₋₁₂ -N ₃₉ -PR / 4	10/10	6	(Edelmann et al., 2002)
DXS7423	12-18	173-204	PF-N ₃₀ -(TCCA) ₃ -TCTGTCCT-(TCCA) ₉₋₁₅ - N ₄₇ -PR ou (TCCA) ₁₂₋₁₈ / 4	14/15	14	(Edelmann et al., 2002)
DXS6809	27-38	235-279 ou 229-280	PF-N ₃₇ -(CTAT) ₇₋₁₀ -(ATCT) ₃ -N ₉ -(TATC) ₃ - (ATCT) _{11,13-17} -N ₃₁ -PR / 4	31/34	31	(Gomes et al., 2007)
DXS7132	9, 11- 12, 16.3, 17.3, 18.3(cat=1)	131-155 ou 130-162	PF-(CTAT) _{9,14-16} (CAT) _{0,2} -PR ou PF-N ₁₉ - (TCTA) ₁₁₋₁₇ -N ₂₈ -PR / 4 ou 4+0.1(CAT)	12/12	13	(I. Gomes, Prinz, et al., 2007)
DXS9902	9-14, 11.1, 12.1, 13.1	120-160 ou 179-231	PF-(TAGA) _n -PF; PF-A _{0,1} -N ₁₂ -(TAGA) ₁₁₋₁₃ - PF / 4 ou 4+0.1(A)	12/12	13	(Leonor Gusmão et al., 2009)
DXS6789	13-25	240-270 ou 249-297	PF-(TATC) _{0,1} -(TATG) ₇₋₁₁ -(TATC) _{6,7,9-10-13} - PR / 4	21/22	20	(I. Gomes, Prinz, et al., 2007)

Nota: PF, número de nucleótidos/bases que compõem o forward primer; N, número de nucleótidos intercalados; PR, número de nucleótidos que compõem o reverse primer; A, designa o nucleótido contendo Adenina.

Os resultados para os três STRs comuns aos dois multiplexes (DXS8378:Xp22.31; DXS7132:centrómeroXq11.2; DXS7423:Xq28) foram iguais, conforme esperado.

4.1 Diversidade Genética

O cálculo da frequência alélica e diversidade genética (Tabela 6 e Tabela 7) foi efetuado com o software Arlequin suite ver 3.5 (Excoffier & Lischer, 2010).

As frequências alélicas e valores de diversidade genética (H_{ESP}) X-STR Decaplex, constam na Tabela 6, e o Investigator Argus X-12, constam da Tabela 7.

*Tabela 6: Frequências alélicas e valores de diversidade genética para os 10 loci STR do X-STR Decaplex-ISFG, numa amostra de Timor-Leste.
(n= 45 cromossomas X, oriundos de 35M:5F)*

Alelos	Grupo de Ligação 1				Grupo de Ligação 2				Grupo de Ligação 4	Grupo de Ligação 4
	DXS837 8	DXS990 2	DXS713 2	DXS989 8	DXS680 9	DXS678 9	DXS713 3	GATA172D 05	GATA31E08	DXS7423
6								0,0667		
8								0,0667		
9	0,0667			0,2667			0,7333	0,0222	0,0444	
10	0,4667						0,2000	0,4000	0,0667	
11	0,1778	0,6000		0,0667			0,0444	0,2889	0,2889	
12	0,2889	0,1111	0,2000	0,4222				0,1333	0,2667	
13		0,2667	0,2000	0,1556			0,2222	0,2222	0,2667	
14		0,0222	0,2444	0,0667					0,0667	0,4222
15			0,2889	0,0222		0,3556				0,4000
16			0,0667			0,3111				0,1556
17										0,0222
20						0,2222				
21						0,0444				
22						0,0444				
23						0,0222				
29					0,0222					
30					0,0889					
32					0,0667					
33					0,1778					
34					0,4889					
35					0,1333					
36					0,0222					
N.ºalelos	5	4	5	6	7	6	4	7	6	4
N.º médio de alelos ±s.d.	5,400±1,174									
H_{ESP}	0,6899	0,56869	0,7899	0,73333	0,71414	0,73939	0,42929	0,74545	0,7808	0,6515
valor médio de HESP ±s.d.	0,68424±0,11057									

Nota:s.d., standard deviation; o mesmo que, dp, desvio padrão da amostra.

Tabela 7: Frequências alélicas e valores de diversidade genética para os 12 loci STR do Investigator® Argus X-12, numa amostra de Timor-Leste. (n= 182 cromossomas X:134 indivíduos)

Alleles	Grupo de Ligação 1			Grupo de Ligação 2			Grupo de Ligação 3			Grupo de Ligação 4		
	DXS1014 8	DXS1013 5	DXS837 8	DXS713 2	DXS1007 9	DXS1007 4	DXS1010 3	HPRTB	DXS1010 1	DXS1014 6	DXS1013 4	DXS742 3
8						0,15385						
9			0,03297			0,01099						
10			0,48901									
11			0,23077	0,01099				0,0384 6				
12			0,21978	0,15385				0,1703 3				
13			0,021978	0,25824				0,5549 5				
14				0,21429			0,00549	0,2033 0,0274				0,4011
15			0,00549	0,31868		0,03297		0,0054 7				0,43407
16				0,04396	0,01648	0,17033	0,15934	9				0,13736
17	0,016484	0,010989			0,12088	0,23077	0,03846					0,02747
17.2						0,00549						
18	0,24176	0,01648			0,15934	0,3022	0,24725					
19	0,04945	0,05495			0,24176	0,07692	0,42308					
20		0,11538			0,23077	0,01648	0,10989					
21	0,01648	0,11538			0,18132		0,01099					
22		0,12088			0,04396		0,00549					
23		0,0989			0,00549							
23.1	0,1044											
24		0,07143								0,04396		
24.1	0,08791											
25		0,06593								0,03297		
25.1	0,12088											
26	0,02198	0,04396								0,13187		
26.1	0,17033											
27	0,00549	0,04396							0,00549	0,13187		
27.1	0,04396											
27.2												
27.3									0,01648			
28	0,021978	0,02747							0,06044			
28.1	0,01648									0,14286		
28.2										0,00549		
29	0,04396	0,04396							0,08242			
29.1	0,00549								0,02747	0,18132		
29.2												
30		0,03846							0,08791			
30.1	0,02198								0,0989	0,12637	0,00549	
30.2												
31	0,00549	0,04396							0,16484			
31.1	0,00549								0,10989	0,08242	0,01099	
31.2												
32		0,01099							0,12088			
32.2									0,0989	0,04396	0,00549	
33		0,03297							0,03297			
33.2									0,04945	0,01648	0,03297	
34		0,00549							0,01099			
34.1									0,02747	0,02747	0,03846	
35		0,02198									0,00549	
36		0,00549							0,00549		0,18132	
37		0,01099									0,22527	
38											0,19231	
38.1											0,13736	
39											0,00549	
39.3											0,06044	
40											0,03297	
40.2											0,01648	
40.3										0,00549		
41.2										0,01099	0,02198	
41.3											0,02747	
42.2										0,00549		
43.2										0,01099		
N.ºalelos	18	21	6	6	8	9	8	6	16	16	16	4
N.º médio de alelos +/-s.d.	11,167±5,797											
HESP	0,87542	0,92939	0,66134	0,76431	0,81768	0,79971	0,72473	0,6228 5	0,9079	0,8877	0,8562	0,6346
valor médio de HESP +/-s.d	0,79015±0,10829											

Os resultados obtidos para os parâmetros forenses mais relevantes, para os três *loci* STR comuns, não apresentam diferenças significativas, aliás conforme esperado, contudo poderia haver sempre a possibilidade de a diferença no tamanho das amostras (45 vs 134) provocar um desvio considerável, não é o caso, conforme Tabela 8.

Tabela 8: Comparação dos valores obtidos para os loci comuns nos multiplex em estudo, DXS8378, DXS7132 e DXS7423

	DXS8378		DXS7132		DXS7423	
	Decaplex	Argus X12	Decaplex	Argus X12	Decaplex	Argus X12
H_{ESP}	0,6899 (5)	0,66134 (6)	0,7899 (5)	0,76431 (6)	0,63457 (4)	0,6346 (4)
Polymorphism Information content (PIC)	0,6044	0,60290	0,734276	0,7200	0,5633	0,5567
PD female	0,8279	0,82803	0,9101	0,9023	0,7945	0,7895
PD male	0,6627	0,65771	0,7723	0,7604	0,6371	0,6311

Nota: Não diferem significativamente. As possíveis diferenças devem-se às diferenças de tamanho da amostra.

Efetuar-se testes de comparação de variância, F de Fisher, e testes de comparação de médias, t de Student.

O teste F de Fisher permitiu concluir que as variâncias são iguais entre si para cada variável.

O teste t de Student pressupõe que a hipótese nula, H_0 , é tal que a média do valor obtido pelo X-Decaplex é igual à média obtida com o Argus X-12; por oposição a hipótese alternativa, H_1 , é tal que a média X-Decaplex \neq média Argus X-12.

Os resultados obtidos foram os seguintes:

Tabela 9: Comparação Heterozigosidade ou diversidade genética para os 3 STRs comuns.

Teste T: duas amostras com variâncias iguais

	<i>HESP X-Decaplex</i>	<i>HESP Argus X-12</i>
Média	0,70479	0,68675
Variância	0,006198136	0,004690422
Observações	3	3
Variância agrupada	0,005444279	
Hipótese de diferença de média	0	
gl	4	
Stat t	0,299441485	
P(T<=t) uni-caudal	0,389758845	
t crítico uni-caudal	2,131846786	
P(T<=t) bi-caudal	0,77951769	
t crítico bi-caudal	2,776445105	

Os valores de $p > 0.05$ e P tende para 1.

O valor $\text{Stat} < t$ crítico, logo os valores caem na zona de aceitação da distribuição normal, sendo a H_0 verdadeira, isto é, para aqueles 3 STRs Hesp real não difere, quer se utilize com o X-Decaplex, quer se utilize o Argus X-12, para um nível de significância de 5 %.

Tabela 10: Comparação do PIC para os 3 STRs comuns.

Teste T: duas amostras com variâncias iguais

	<i>PIC X-Decaplex</i>	<i>PIC Argus X-12</i>
Média	0,633992	0,626533333
Variância	0,007964963	0,007085623
Observações	3	3
Variância agrupada	0,007525293	
Hipótese de diferença de média	0	
gl	4	
Stat t	0,10530406	
P(T<=t) uni-caudal	0,46060194	
t crítico uni-caudal	2,131846786	
P(T<=t) bi-caudal	0,92120388	
t crítico bi-caudal	2,776445105	

Tabela 11: Comparação do PD female para os 3 STRs comuns.

Teste T: duas amostras com variâncias iguais

	PD female X-Decaplex	PD female Argus X-12
Média	0,844166667	0,839943333
Variância	0,003539293	0,003287406
Observações	3	3
Variância agrupada	0,003413349	
Hipótese de diferença de média	0	
gl	4	
Stat t	0,088534107	
P(T<=t) uni-caudal	0,466853814	
t crítico uni-caudal	2,131846786	
P(T<=t) bi-caudal	0,933707628	
t crítico bi-caudal	2,776445105	

Para o PIC e PDfemale também não existe diferença significativa entre os resultados para os 3 STRs comuns ao X-Decaplex e Argus X-12, conforme se pode confirmar pelos valores de probabilidade (P) e t calculado (Stat).

Tabela 12: Comparação do PDmale para os 3 STRs comuns.

Teste T: duas amostras com variâncias iguais

	PD male X-Decaplex	PD male Argus X-12
Média	0,6907	0,68307
Variância	0,00515776	0,00466197
Observações	3	3
Variância agrupada	0,004909865	
Hipótese de diferença de média	0	
gl	4	
Stat t	0,133363013	
P(T<=t) uni-caudal	0,450173318	
t crítico uni-caudal	2,131846786	
P(T<=t) bi-caudal	0,900346636	
t crítico bi-caudal	2,776445105	

Igual conclusão pode tirar-se para o parâmetro PDmale..

Logo H_{esp}, PIC, PD_{female} e PD_{male}, não diferem, quer sejam obtidos com recurso ao X-Decaplex ou pelo Argus X-12, independentemente do tamanho das amostras, conforme demonstrado.

4.2 Comparação de Métodos

Em relação à diferença entre as médias referentes à diversidade genética entre multiplexs, pode concluir-se que o multiplex Argus é mais potente, pois oferece uma maior amplitude de alelos que permitem diferenciar melhor os perfis entre si.

A diversidade genética é significativamente maior ($H_{ESP}^{ArgusX12} > H_{ESP}^{XSTRDECA}$, ver Tabela 13) logo oferece um poder discriminatório superior.

Tabela 13: Comparação entre H_{ESP} obtidos para cada multiplex.

	Média	
	Decaplex	Argus X12
H_{ESP} (DG)	0,6842±0,1106	0,7902±0,1083
N.º alelos	5,400±1,174	11,167±5,797

Tabela 14: Comparação entre a Diversidade Genética obtida para o Argus X-12 e X-Decaplex

Teste T: duas amostras com variâncias desiguais

	H_{ESP} (Argus X-12)	H_{ESP} (X-Decaplex)
Média	0,790150833	0,684242
Variância	0,011726316	0,012226196
Observações	12	10
Hipótese de diferença de média	0	
gl	19	
Stat t	2,258080071	
P(T<=t) uni-caudal	0,017949186	
t crítico uni-caudal	1,729132812	
P(T<=t) bi-caudal	0,035898372	
t crítico bi-caudal	2,093024054	

Recordando, a hipótese nula afirma que as médias da Diversidade Genética, DG, obtida com o Argus X-12 é igual à média com o X-Decaplex e a hipótese alternativa diz o oposto.

O valor de p tende para zero e é inferior a 0.05 e t calculado ou Stat > t crítico, logo cai na área de rejeição da hipótese nula, por isso, podemos afirmar que os valores de diversidade genética obtidos são diferentes entre si, com um grau de confiança de 95% e com nível de significância de 5%.

A mesma conclusão se repete quando se comparam os restantes parâmetros forenses.

4.3 Estatística populacional para efeitos forenses.

Os restantes parâmetros de estatística populacional para fins forenses (Tabela 45 e Tabela 16) foram, por mim, calculados através da aplicação online no site internacional do CrX, Forensic ChrX Research (ChrX-STR-Org, 2014). A calculadora encontra-se selecionando no painel de menus lateral, o menu EVALUATION & CALCULATE e, de seguida, no menu CALCULATE, inserir os dados (alelo vs frequência obtidos no Arlequin) de um apenas um *loci*, devendo refrescar a página a cada introdução de dados de outro *loci*.

Os parâmetros forenses são a medida que permite avaliar se os *loci* são adequados para usar na investigação forense.

Um *locus* cujo conteúdo informativo polimórfico e/ou poder discriminatório (PIC e PD) acima de 0,5 é considerado um marcador muito informativo. No sistema X-STR Decaplex encontramos dois marcadores pouco informativos, o DXS9902 e o DXS7133, assinalados a bold. Este último *locus* tem vindo a cair em desuso de acordo com a literatura científica atual e mais uma vez os dados experimentais abaixo descritos convergem para a mesma conclusão. O locus mais informativo é o DXS7132, assinalado com um asterisco, apresenta o valor de 0.7343.

Os 12 *loci* caracterizados com o *Investigator* Argus X-12, todos apresentam um conteúdo polimórfico informativo acima de 0.5, logo todos os *loci* têm interesse para investigação forense.

O *locus* DXS7423, em comum com o Decaplex, apresenta o valor mais baixo, 0.55672, o *locus* mais informativo é DXS10135, com 0.91929, o qual está em sintonia com o total de alelos que evidenciou. Entre todos os STRs estudados, DXS10135 apresentou o maior número de alelos, foram identificados 21 alelos para este *locus*.

Tabela 15: Parâmetros de estatística populacional com interesse forense para 10 loci STR numa amostra de Timor-Leste. (n= 45 cromossomas X; 40 indivíduos)

	Grupo de Ligação 1				Grupo de Ligação 2			Grupo de Ligação 3		Grupo de Ligação 4
	DXS 8378	DXS9902	DXS7132	DXS 9898	DXS6809	DXS6789	DXS 7133	GATA172D05	GATA31E08	DXS7423
Polymorphism Information content (PIC):	0,6044	0,4938	0,7343*	0,6744	0,6658	0,6740	0,3739	0,6883	0,7246	0,5633
Homozygosity (h):	0,3373	0,4439	0,2277	0,2830	0,3017	0,2770	0,5802	0,2711	0,2366	0,3629
Heterozygosity (HET):	0,6627	0,5561	0,7723	0,7170	0,6983	0,7230	0,4198	0,7289	0,7634	0,6371
Power of Exclusion (PE):	0,3730	0,2415	0,5488	0,4551	0,4256	0,4647	0,1264	0,4743	0,5330	0,3773
Paternity Index (PI):	0,1686	0,2220	0,1138	0,1415	0,1509	0,1385	0,2901	0,1356	0,1183	0,1815

Power of Discrimination										
PD female:	0,8279	0,7407	0,9101	0,8773	0,8764	0,8743	0,6174	0,8859	0,9052	0,7945
PD male:	0,6627	0,5561	0,7723	0,7170	0,6983	0,7230	0,4198	0,7289	0,7634	0,6371

Mean paternity exclusion chance										
MEC Krüger:	0,4007	0,3013	0,5506	0,4866	0,4868	0,4785	0,2136	0,5049	0,5423	0,3541
MEC Kishida:	0,6044	0,4937	0,7343	0,6744	0,6658	0,6740	0,3739	0,6883	0,7246	0,5632
MEC Desmarais:	0,6044	0,4938	0,7343	0,6744	0,6658	0,6740	0,3739	0,6883	0,7246	0,5633
MEC Desmarais Duo:	0,4571	0,3485	0,5997	0,5327	0,5224	0,5321	0,2420	0,5489	0,5896	0,4172

Tabela 16: Parâmetros de estatística populacional com interesse forense para os 12 loci STR do Investigator® Argus X-12, numa amostra de Timor-Leste. (n= 182 cromossomas X; 134 indivíduos)

	Grupo de Ligação 1			Grupo de Ligação 2			Grupo de Ligação 3			Grupo de Ligação 4		
	DXS1014 8	DXS1013 5	DXS837 8	DXS713 2	DXS1007 9	DXS1007 4	DXS1010 3	HPRT B	DXS1010 1	DXS1014 6	DXS1013 4	DXS742 3
Polymorphism Information content (PIC):	0,85853	0,91929	0,60290	0,72000	0,78648	0,76603	0,67935	0,5719 8	0,89490	0,87155	0,83481	0,55672
Homozygosity (h):	0,12940	0,07572	0,34229	0,23989	0,18682	0,20469	0,27924	0,3805 8	0,09709	0,11714	0,14853	0,36892
Heterozygosity (HET):	0,87061	0,92429	0,65771	0,76011	0,81319	0,79531	0,72076	0,6194 2	0,90291	0,88286	0,85147	0,63108
Power of Exclusion (PE):	0,73586	0,84525	0,36592	0,52723	0,62374	0,59037	0,46110	0,3148 4	0,80137	0,76056	0,69776	0,32985
Paternity Index (PI):	0,06470	0,03786	0,17115	0,11994	0,09341	0,10234	0,13962	0,1902 9	0,04959	0,05857	0,07427	0,18446

Power of Discrimination												
PD female:	0,97118	0,98927	0,82803	0,90234	0,93840	0,92882	0,88062	0,8077 2	0,98256	0,97497	0,96128	0,78954
PD male:	0,87061	0,92429	0,65771	0,76011	0,81319	0,79531	0,72076	0,6194 2	0,90291	0,88286	0,85147	0,63108

Mean paternity exclusion chance												
MEC Krüger:	0,74592	0,84704	0,40195	0,53368	0,62728	0,60088	0,49179	0,3793 8	0,80364	0,76399	0,70605	0,34967
MEC Kishida:	0,85852	0,91929	0,60290	0,72000	0,78648	0,76603	0,67931	0,5719 8	0,89489	0,87155	0,83478	0,55672
MEC Desmarais:	0,85853	0,91929	0,60290	0,72000	0,78648	0,76603	0,67935	0,5719 8	0,89490	0,87155	0,83481	0,55672
MEC Desmarais Duo:	0,76420	0,85567	0,45531	0,58338	0,66465	0,63959	0,53779	0,4231 8	0,81699	0,78200	0,73042	0,41109

Capítulo VI Conclusão.

5 Discussão e Conclusão.

Este trabalho permitiu caracterizar de uma forma simples a população de Timor Leste, por aplicação de dois sistemas de estudo de STRs do cromossoma, o *X-STR Decaplex*-GHEP-ISFG e o do Investigator® Argus X-12.

O sistema multiplex X-STR Decaplex, desenvolvido pelo GHEP-ISFG, é uma das ferramentas mais utilizadas em estudos populacionais sempre com o objetivo de tornar mais frequente o seu uso como uma ferramenta para fins forenses. A adesão ao multiplex teve uma trajetória exponencial entre 2007 e 2013, tendo sido muito bem aceite entre a comunidade científica, um pouco por todo o mundo. O desenvolvimento do multiplex da Qiagen Argus tem sofrido uma evolução substancial na sua composição, tendo sido acrescentados vários STR a genotipar ao longo das etapas de remodelação do kit. O maior número de STRs e sua variabilidade aumentaram a certeza na tomada de decisões pois permite recolher mais informação a partir da mesma amostra, permitindo descrevê-la melhor. Este multiplex passou do estudo de 6 para 12 STR, entre os quais estão os que apresentam maior diversidade genética.

O Decaplex também tem vindo a ser otimizado, mas de uma forma mais lenta e não sistematizada, de tal forma que muitos são os que utilizam o atual Argus X-12, sobretudo as instituições forenses, pois estão bem estudados e cumprem com os preceitos das normas de qualidade, sendo fácil validar o seu uso em laboratório. Ao serem comercializados na forma *kits* permitem poupar tempo, sistematizar e otimizar o trabalho, pois o *kit* fornece todas as soluções necessárias, incluindo o *ladder* alélico, DNA de referência, informação necessária para a execução prática e referências teóricas.

O poder de discriminação (PD) de género é outro dos parâmetros estatísticos que permitem avaliar a eficiência ou informação de um *loci* em mulheres PD_{female} (PD_f) ou em homens PD_{male} (PD_m).

O PD é outro parâmetro que permite escolher os *loci* mais eficazes na aplicação forense.

O X-Decaplex, apesar de ser mais trabalhoso, tem estado na origem de diversas publicações e inúmeros artigos.

A tabela que se segue reúne um conjunto de dados obtidos em populações com interesse e/ou afinidade com a população de Timor.

Tabela 17: Comparação entre populações para resultados obtidos após estudo com os 10 loci do X-STR Decaplex-ISFG- PD_{female}

	DXS8378		DXS9902		DXS7132		DXS9898		DXS6809		DXS6789		DXS7133		GATA172D05		GATA31E08		DXS7423	
	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m	PD _f	PD _m
Timor-Leste	0,827	0,662	0,740	0,556	0,910	0,772	0,877	0,717	0,876	0,698	0,874	0,723	0,617	0,419	0,885	0,728	0,905	0,763	0,794	0,637
Macau	0,958	0,837	0,958	0,837	0,886	0,726	0,849	0,681	0,994	0,939	0,994	0,939	0,499	0,317	0,904	0,749	0,925	0,789	0,678	0,536
Nicarágua	0,825	0,664	0,825	0,667	0,910	0,766	0,872	0,708	0,917	0,774	0,910	0,766	0,745	0,549	0,933	0,803	0,901	0,741	0,853	0,693
Colla (Arg)	0,781	0,625	0,691	0,502	0,852	0,697	0,742	0,573	0,891	0,725	0,718	0,516	0,73	0,566	0,875	0,709	0,781	0,587	0,811	0,635
Toba (Arg)	0,777	0,622	0,705	0,499	0,919	0,776	0,826	0,67	0,824	0,619	0,697	0,494	0,657	0,461	0,738	0,561	0,572	0,36	0,776	0,596
Angola	0,857	0,713			0,885	0,742	0,926	0,791	0,957	0,856	0,955	0,853			0,881	0,734			0,845	0,684
Moçambique	0,838	0,694			0,890	0,743	0,935	0,812	0,971	0,88	0,962	0,866			0,916	0,783			0,791	0,623
Uganda	0,827	0,684			0,889	0,751	0,943	0,836	0,943	0,8329	0,948	0,841			0,930	0,814			0,860	0,709
Colômbia	0,755	0,6489	0,7829	0,6697	0,8645	0,7393	0,9189	0,7985	0,9236	0,8045	0,9168	0,7959	0,7687	0,6589	0,9391	0,8255	0,9011	0,7773	0,8141	0,6944
Marrocos	0,847191	0,69799	0,86999	0,715043	0,90222	0,757877	0,917118	0,773397	0,955924	0,84222	0,949653	0,827622	0,864751	0,709834	0,940005	0,815855	0,932144	0,803429	0,835086	0,673943
Perú	0,8055	0,6389	0,7946	0,6231	0,9267	0,7913	0,8417	0,6783	0,9207	0,7759	0,8803	0,7025	0,7265	0,5372	0,9038	0,7542	0,8639	0,6944	0,8637	0,7026

Dados populacionais expostos na tabela correspondem à seguintes fontes de informação, Timor-Leste, este estudo; Macau, (I. Gomes, Amorim, Pereira, Carracedo, & Gusmão, 2009); Nicarágua, (Núñez et al., 2013); Argentina (Colla,Toba), (Toscanini, Gusmão, Berardi, & Raimondi, 2009); Angola, Moçambique e Uganda (I. Gomes, Alves, et al., 2007); Colômbia, (Martínez et al., 2011); Marrocos, (Laouina, Aznar, Chbel, & de Pancorbo, 2013) e Perú (Barletta et al., 2013). Estas populações contêm os dados mais recentes para o fim que se pretende, contribuindo para manter a atualidade e utilidade do multiplex.

A cor-de-rosa, estão representados os valores iguais, pois os grupos DXS8378/DXS9902 e DXS6809/DXS6789 são frequentemente representados como haplótipos por se recombinarem em bloco, como se de um só *loci* se tratasse. Na Tabela 17 não estão representados os valores de alguns *loci*. Por norma, só se utilizam os *loci* mais informativos ou polimórficos, pelo que se depreende que os *loci* em falta não foram considerados úteis para aqueles estudos. Através da consulta da respetiva literatura, pode verificar-se que foram substituídos por outros *loci*.

Todos os *loci* são bastante informativos mas verifica-se uma tendência geral para o serem menos no caso de perfis masculinos e para cada *locus*.

Os *loci* com menor poder discriminatório são DXS9902, DXS7133 e o DXS7423.

Os mais informativos são DXS6809, DXS6789, DXS7132; este último é o único em comum com o conjunto de *loci* do *Investigator* Argus X-12.

Na amostra da população de Timor-Leste os *loci* com maior poder discriminatório são: DXS7132, GATA172D05, GATA31E08.

A comparação entre o *Investigator* Argus X12 e o X-STR Decaplex-GHEP ISFG permite concluir que o Argus é muito mais informativo, suficiente e consistente. A caracterização com o X-STR Decaplex-GHEP ISFG pode ainda ser otimizada. Inicialmente, deve ser mais aprofundada, repetindo o estudo para uma amostra maior e incluindo um maior número de indivíduos do sexo feminino, de forma a representar a população a ser analisada. Para além disso, da aplicação do X-STR *Decaplex*, conclui-se que se deve implementar o uso de um *ladder* alélico produzido, por exemplo, a partir dos fragmentos obtidos durante a amplificação, e após a sua separação em gel de agarose. É uma tarefa morosa mas que de futuro abreviaria a classificação alélica dos perfis. De notar também que a literatura relativa ao procedimento de genotipagem dos 10 *loci* STR do CrX revela o recurso a *primers* de diferente composição sequencial para os mesmos *loci*, podendo estes ser uma alternativa a alguns dos *primers* utilizados neste trabalho, visto verificar-se alguma dificuldade em amplificar algumas amostras.

O futuro do X-STR Decaplex deve evoluir no sentido da inclusão de novos *loci*, acautelando os grupos de ligação mas, sobretudo e de certeza, na exclusão dos *loci* menos informativos, como é o caso do DXS7133 e do DXS9902, sob pena de cair em desuso. Como conclusão, este processo de estudo do multiplex do GHEP-ISGF deve conservar a sua originalidade de forma a manter-se como uma alternativa válida e complementar ao sistema Argus X-12.

"We shall not cease from exploration. And the end of all our exploring will be to arrive where we started, and know the place the first time."

T. S. Eliot

Capítulo VII Bibliografía

- Allen, G. E. (2008a). Thomas Hunt Morgan-Complete Dictionary of Scientific Biography. Retrieved April 23, 2014, from http://www.encyclopedia.com/topic/Thomas_Hunt_Morgan.aspx
- Allen, G. E. (2008b, March 5). Clarence Erwin McClung-Complete Dictionary of Scientific Biography. Retrieved from <http://www.encyclopedia.com/doc/1G2-2830902724.html>
- Balderman, S., & Lichtman, M. a. (2011). A history of the discovery of random x chromosome inactivation in the human female and its significance. *Rambam Maimonides Medical Journal*, 2(3), e0058. doi:10.5041/RMMJ.10058
- Barletta, C., Builes, J. J., Bravo, M. L., Aguirre, D., Descailleaux, J., & Gusmão, L. (2013). Genetic data of 10 X-STRs in a population sample from Lima, Perú. *Forensic Science International: Genetics Supplement Series*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1875176813000887>
- Barr, M. L., & Bertram, E. G. (1949). A morphological distinction between neurones of the male and female, and the behavior of the nucleolar satellite during accelerated nucleoprotein synthesis., 163, 676–677.
- Bennett, P. (2000). Microsatellites. *Journal of Clinical Pathology: Molecular Pathology*, 53(4), 177–183. doi:10.1136/mp.53.4.177
- Berger, J. D. (2008). Henking, Hermann - Complete Dictionary of Scientific Biography. Retrieved March 24, 2014, from <http://www.encyclopedia.com/doc/1G2-2830901934.html>
- BGBX. (2014). Banco Genético Brasileiro do Cromossomo X. *Banco Genético Brasileiro do Cromossomo X*. Retrieved October 27, 2014, from www.bgbx.com.br
- Biochemical Nomenclature Committees, I. & I. (1974). INTERNATIONAL UNION OF BIOCHEMISTRY ABBREVIATIONS AND SYMBOLS FOR NUCLEIC ACIDS , POLYNUCLEOTIDES AND THEIR CONSTITUENTS Rules Approved 1974. Retrieved from <http://www.chem.qmul.ac.uk/iupac/misc/naabb.html#p131>
- Boumil, R. M., & Lee, J. T. (2001). Forty years of decoding the silence in X-chromosome inactivation. *Human Molecular Genetics*, 10(20), 2225–32. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11673405>

- Brenner, C. H. (2015). DNA profile probability. Retrieved September 18, 2015, from <http://dna-view.com/profile.htm>
- Butler, J. M. (2012). *Advanced Topics in Forensic DNA Typing*. (A. Press, Ed.) *Advanced Topics in Forensic DNA Typing* (3rd ed.). AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK OXFORD • PARIS • SAN DIEGO • SAN FRANCISCO SINGAPORE • SYDNEY • TOKYO: Elsevier Ltd. doi:10.1016/B978-0-12-374513-2.00009-9
- Carey, F. A. (1992). *Organic Chemistry*. (I. McGraw-Hill, Ed.) (2nd ed.). New York; St Louis; San Francisco; Auckland; Bogotá; Caracas; Lisbon; London; Madrid; Mexico; Milan; Montreal; New Delhi; Paris; San Juan; Singapore; Sydney; Tokyo; Toronto: McGraw-Hill, Inc.
- Carlson, E. A. (2013). *The 7 Sexes: Biology of Sex Determination*. Indiana University Press. Retrieved from https://books.google.pt/books?id=viVlcBwtqkcC&pg=PA146&hl=pt-PT&source=gbs_toc_r&cad=3#v=onepage&q&f=false
- Center for Regional Studies, U. K. A. W. (1997). *O mapa linguístico de Timor Leste*. Ethnologue, Language data from Center for Regional Studies, Universitas Kristen Artha Wacana.
- ChrX-STR-Org. (2014). Forensic ChrX Research. *ChrX-STR.org 2.0*. Retrieved October 27, 2014, from chrX-str.org
- Commons, W. (2014). DNA-Base Pairing. *Wikimedia Commons*. Retrieved October 6, 2014, from http://en.wikipedia.org/wiki/DNA#Base_pairing
- Darryl Leja. (2010). NHGRI-85154. NHGRI. Retrieved from <http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85154>
- Dashnow, H., Tan, S., Das, D., Easteal, S., & Oshlack, A. (2015). Genotyping microsatellites in next-generation sequencing data. *BMC Bioinformatics*, 16(Suppl 2), A5. doi:10.1186/1471-2105-16-S2-A5
- Dennell, R., & Petraglia, M. D. (2012). The dispersal of Homo sapiens across southern Asia: how early, how often, how complex? *Quaternary Science Reviews*, 47, 15–22. doi:10.1016/j.quascirev.2012.05.002
- DOE-US. (2006). IHG Sequencing Centers. Retrieved September 9, 2015, from <https://www.genome.gov/11006939>
- DOE-US. (2013). Human Genome Project Research Sites. *U.S. Department of Energy Human Genome Project*. Retrieved September 9, 2015, from http://web.ornl.gov/sci/techresources/Human_Genome/research/centers.shtml

- Dose, K. (1982). *Bioquímica*. (E.-E. P. e U. E. S.-V. Lda, Ed.) (1a Edição.). São Paulo, Brasil: EPU- Editora Pedagógica e Universitária.
- Edelmann, J., Deichsel, D., Hering, S., Plate, I., & Szibor, R. (2002). Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377. *Forensic Science International*, 129(2), 99–103. doi:S037907380200230X [pii] ET - 2002/09/24
- Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics*, 5(6), 435–445. doi:10.1038/nrg1348
- Ensembl release. (2015a). Chromosome X- Ensembl summary. *Ensembl release* 81. Retrieved January 1, 2015, from http://www.ensembl.org/Homo_sapiens/Location/Chromosome?r=X
- Ensembl release. (2015b). Chromosome Y- Ensembl summary. *Ensembl release* 81. Retrieved September 21, 2015, from http://www.ensembl.org/Homo_sapiens/Location/Chromosome?r=Y:32994743-33094743
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Fernández, M. C. (2013). Estudio de los microsatélites y miniSTRs del cromosoma X de aplicación forense.
- GHEP-ISFG. (1995). Retrieved October 25, 2015, from <http://www.gep-isfg.org/pt/>
- Gomes, C., Magalhães, M., Alves, C., Amorim, A., Pinto, N., & Gusmão, L. (2012). Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. *International Journal of Legal Medicine*, 126(6), 917–21. doi:10.1007/s00414-012-0768-5
- Gomes, I., Alves, C., Maxzud, K., Pereira, R., Prata, M. J., Sánchez-Diz, P., ... Gusmão, L. (2007). Analysis of 10 X-STRs in three African populations. *Forensic Science International. Genetics*, 1(2), 208–11. doi:10.1016/j.fsigen.2007.01.001
- Gomes, I., Amorim, A., Pereira, V., Carracedo, A., & Gusmão, L. (2009). Genetic patterns of 10 X chromosome short tandem repeats in an Asian population from Macau. *Forensic Science International: Genetics Supplement Series*, 2(1), 402–404. doi:10.1016/j.fsigss.2009.09.009

- Gomes, I., Prinz, M., Pereira, R., Meyers, C., Mikulasovich, R. S., Amorim, A., ... Gusmão, L. (2007). Genetic analysis of three US population groups using an X-chromosomal STR decaplex. *International Journal of Legal Medicine*, 121(3), 198–203. doi:10.1007/s00414-006-0146-2
- Gomes, S. M., Bodner, M., Souto, L., Zimmermann, B., Huber, G., Strobl, C., ... Parson, W. (2015). Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocenic mtDNA diversity. *BMC Genomics*, 16, 6–12. doi:10.1186/s12864-014-1201-x
- Goudet, J. (1993). *The Genetics of Geographically Structured Populations*. University of Wales, College of Bangor. Retrieved from www2.unil.ch/popgen/research/reprints
- Graves, J. A. M. (2006). Sex chromosome specialization and degeneration in mammals. *Cell*, 124(5), 901–14. doi:10.1016/j.cell.2006.02.024
- Graves, J. A. M. (2010). Review: Sex chromosome evolution and the expression of sex-specific genes in the placenta. *Placenta*, 31 Suppl, S27–32. doi:10.1016/j.placenta.2009.12.029
- Gusmão, L., Alves, C., Sánchez-Diz, P., Zarrabeitia, M. T., Abovich, M. a., Aragón, I., ... Amorim, a. (2008). Results of the GEP-ISFG collaborative study on an X-STR Decaplex. *Forensic Science International: Genetics Supplement Series*, 1(1), 677–679. doi:10.1016/j.fsigss.2007.10.012
- Gusmão, L., Sánchez-Diz, P., Alves, C., Gomes, I., Zarrabeitia, M. T., Abovich, M., ... Amorim, A. (2009). A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *International Journal of Legal Medicine*, 123(3), 227–34. doi:10.1007/s00414-008-0309-4
- Hardy, G. H. (2003). Mendelian proportions in a mixed population. 1908. *The Yale Journal of Biology and Medicine*, 76(2), 79–80. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2582692&tool=pmc&rendertype=abstract>
- Henking, H. (1891). Über Spermatogenese und deren Beziehung zur Eientwicklung bei *Pyrrhocoris apterus* L. *Zeit schrift für wissenschaftliche. Zoologie*, (51), 685–736.
- Human Genome Project Information Archive - About the Human Genome Project. (2014). *U.S. DOE Human Genome Project*. Retrieved October 12, 2014, from http://web.ornl.gov/sci/techresources/Human_Genome/project/index.shtml
- Ilona Miko. (2008). Sex Chromossomes and Sex Determination. *Nature Education*,

- 1(1), 108. Retrieved from <http://www.nature.com/scitable/topicpage/sex-chromosomes-and-sex-determination-44565>
- Jeon, Y., Sarma, K., & Lee, J. T. (2012). New and Xisting regulatory mechanisms of X chromosome inactivation. *Current Opinion in Genetics & Development*, 22(2), 62–71. doi:10.1016/j.gde.2012.02.007
- Kobachuk, L. d'Avila G. (2012). *Estudo de Frequências Alélicas de Dez Locos STRs do Cromossomo X na População do Estado do Paraná e sua Contribuição na Identificação Humana*. Universidade Federal do Paraná.
- Lahn, B. T., & Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science*. Retrieved from <http://dx.doi.org/10.1126/science.286.5441.964>
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. doi:10.1038/35057062
- Laouina, A., Aznar, J. M., Chbel, F., & de Pancorbo, M. M. (2013). Genetic data of X-chromosomal STRs in a Moroccan population sample (Casablanca) using an X-STR decaplex. *Forensic Science International: Genetics*, 2–4. doi:10.1016/j.fsigen.2013.02.006
- Lee, J. T., & Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*, 152(6), 1308–23. doi:10.1016/j.cell.2013.02.016
- Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions, and mutational mechanism: a review. *Molecular Ecology*, 11, 253–256.
- Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.) . *Nature* .
- Martínez, B., Builes, J. J., Gusmão, L., Manrique, M., Aguirre, D., Puerto, Y., ... Bravo, M. L. (2011). Genetic data of 10 X-STR in a Colombian population of Bolivar Department. *Forensic Science International: Genetics Supplement Series*, 3(1), e59–e60. doi:10.1016/j.fsigss.2011.08.029
- McClung, C. (1901). Notes on the accessory chromosome. *Anatomischer Anzeiger*, 220–226. Retrieved from <http://www.esp.org/foundations/genetics/classical/cem-01.pdf>
- McClung, C. (1902). The accessory chromosome—Sex determinant? *Biological Bulletin*, (3), 43–44. Retrieved from <http://www.esp.org/foundations/genetics/classical/cem-01.pdf>

- Moreira, H., Costa, H., Tavares, F., & Souto, L. (2015). Genetic variation of 12 X-chromosomal STR loci in an East Timor sample. *International Journal of Legal Medicine*, 129, 257–258. doi:10.1007/s00414-014-1126-6
- Morgan, T. H. (1909). What are “Factors” in Mendelian Explanations? *American Breeder Association Reports*, (5), 365–368. Retrieved from <http://www.esp.org/foundations/genetics/classical/thm-10a.pdf>
- Morgan, T. H. (1910). Sex Limited Inheritance in *Drosophila melanogaster*. *Science*, (32), 120–122. Retrieved from <http://www.esp.org/foundations/genetics/classical/thm-10a.pdf>
- Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods in Enzymology*, 155(155), 335–350.
- NCBI. (2015). X-Chromosome. Retrieved October 28, 2015, from <http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?ORG=hum&MAPS=ideogr,est,loc&LINKS=ON&VERBOSE=ON&CHR=X>
- Nora, E. P., & Heard, E. (2009). X chromosome inactivation: when dosage counts. *Cell*, 139(5), 865–867. doi:10.1016/j.cell.2009.11.009
- Núñez, C., Baeta, M., Aznar, J. M., Sosa, C., Casalod, Y., Bolea, M., ... Martínez Jarreta, B. (2013). Genetic diversity of 10 X chromosome STRs in an admixed population of Nicaragua. *Forensic Science International. Genetics*, 1–2. doi:10.1016/j.fsigen.2013.02.008
- Ohno, S. (1967). Sex Chromosomes and Sex-linked Genes. Retrieved from <http://www.cabdirect.org/abstracts/19680100985.html;jsessionid=D50A2CB0ABE9EA0E00DB81ECC82301E0>
- Oppenheimer, S. (2014). JOURNEY OF MANKIND - The Peopling of the World. *BradShaw Foundation & Stephen Oppenheimer*. Retrieved October 24, 2014, from <http://www.bradshawfoundation.com/journey/>
- Payer, B., & Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics*, 42, 733–772. doi:10.1146/annurev.genet.42.110807.091711
- Payer, B., Lee, J. T., & Namekawa, S. H. (2011). X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells. *Human Genetics*, 130(2), 265–280. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21667284>
- Pessia, E., Engelstädter, J., & Marais, G. a B. (2013). The evolution of X chromosome inactivation in mammals: the demise of Ohno’s hypothesis?

- Cellular and Molecular Life Sciences : CMLS*. doi:10.1007/s00018-013-1499-6
- Qiagen. (2010). Investigator Argus X-12 Handbook, (April). Retrieved from <http://www.qiagen.com/pt/resources/download.aspx?id={5E39E292-638D-4E03-9BF9-142C80106466}&lang=en&ver=1>
- Roque, R. (2008). Histórias de crânios e o problema da classificação antropológica de Timor. *E-Cadernos Ces*, 1(2008), 13–36.
- Ross, M. T., Grafham, D. V, Coffey, A. J., Al, E., Scherer, S., McLay, K., ... Richard Durbin¹, Tim Hubbard¹, Richard A. Gibbs², Stephan Beck¹, J. R. & D. R. B. (2009, April 5). The DNA sequence of the human X chromosome-Author manuscript. *Europe PMC Funders Author Manuscripts*, (April), 1–31. doi:10.1038/nature03440
- Ross, M. T., Grafham, D. V, Coffey, A. J., Scherer, S., McLay, K., Muzny, D., ... Bentley, D. R. (2005). The DNA sequence of the human X chromosome. *Nature*, 434(7031), 325–337. doi:10.1038/nature03440
- Ruiz, Y., Phillips, C., Gomez-Tato, a, Alvarez-Dios, J., Casares de Cal, M., Cruz, R., ... Lareu, M. V. (2013). Further development of forensic eye color predictive tests. *Forensic Science International. Genetics*, 7(1), 28–40. doi:10.1016/j.fsigen.2012.05.009
- Santos, C., Phillips, C., Fondevila, M., Porras-Hurtado, L., Carracedo, A., Souto, L., & Lareu, M. (2011). A study of East Timor variability using the SNPforID 52-plex SNP panel. *Forensic Science International. Genetics*, 5(1), e25–6. doi:10.1016/j.fsigen.2010.03.006
- Schaffner, S. F. (2004). The X chromosome in population genetics. *Nature Reviews. Genetics*, 5(1), 43–51. doi:10.1038/nrg1247
- Sex-Determination System. (2014). *Wikipedia Online - Wikipedia, the free encyclopedia*. Retrieved October 24, 2014, from en.wikipedia.org/wiki/Sex-determination_system
- Silva, R. H. A. da. (2007). *Estudo de Frequência Alélica de Cinco Loci STR do Cromossomo X na População do Estado de São Paulo e sua Contribuição na Identificação Humana*. Universidade de São Paulo.
- Soares, L. M. M. C. V. P. (2014). *Línguas em Timor-Leste : que gestão escolar do plurilinguismo?* Retrieved from <http://ria.ua.pt/handle/10773/12037>
- Souto, L., Alves, C., Gusmão, L., Ferreira, E., Amorim, a, Côrte-Real, F., & Vieira, D. N. (2005). Population data on 15 autosomal STRs in a sample from East Timor. *Forensic Science International*, 155(1), 77–80. doi:10.1016/j.forsciint.2005.01.002

- Souto, L., Gusmão, L., Amorim, A., Corte-real, F., & Vieira, D. N. (2006). Original Research Article Y-STR Haplotype Diversity in Distinct Linguistic Groups From East Timor. *American Journal of Human Biology*, 18(5), 691–701. doi:10.1002/ajhb.2553
- Souto, L., Gusmão, L., Amorim, A., Ferreira, E., Côte-Real, F., Vieira, D. ., & da Cruz e Silva, E. . (2004). POP-Gene TIMOR: first forensic DNA marker study of East-Timor people. *International Congress Series*, 1261, 201–203. doi:10.1016/S0531-5131(03)01533-4
- Souto, L., Gusmão, L., Ferreira, E., Amorim, A., Côte-Real, F., & Vieira, D. N. (2006). Y-chromosome STR haplotypes in East Timor: forensic evaluation and population data. *Forensic Science International*, 156(2-3), 261–5. doi:10.1016/j.forsciint.2005.02.015
- Souto, L., Gusmão, L., Ferreira, E., Pires, A., Rocha, A. M., Amorim, A., ... Vieira, D. N. (2006). Y-chromosome haplotypes in East Timor (Timor-Leste): Evidences of population differentiation. *International Congress Series*, 1288, 256–258. doi:10.1016/j.ics.2005.09.081
- Souto, L., Rocha, a. M., Pires, a., Ferreira, E., Kayser, M., Amorim, a., ... Vieira, D. N. (2006). Mitochondrial DNA variability in populations from East Timor (Timor Leste). *International Congress Series*, 1288, 115–117. doi:10.1016/j.ics.2005.11.060
- Souto, L., Tavares, F., Moreira, H., & Corte-Real, F. (2013). Evaluation of the PowerPlex Fusion System in a sample from East Timor. *Forensic Science International: Genetics Supplement Series*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1875176813000553>
- Spudeit, D. A., Dolzan, M. D., & Micke, G. A. (2012). Conceitos básicos em Eletroforese Capilar. *Scientia Chromatographica*, 4(4), 287–297.
- Technologies, A. (2010). *October 2010 QIAGEN® Multiplex PCR Handbook For fast and efficient multiplex PCR without optimization Sample & Assay Technologies QIAGEN Sample and Assay Technologies*.
- Timor-Leste, G. de. (2010). Sobre Timor - Censos 2010. *Governo de Timor-Leste*. Retrieved September 18, 2015, from timor-leste.gov.tl/?p=547&print=1&lang=pt
- Toscanini, U., Gusmão, L., Berardi, G., & Raimondi, E. (2009). Genetic data of 10 X-STR in two Native American populations of Argentina. *Forensic Science International: Genetics Supplement Series*, 2(1), 405–406. doi:10.1016/j.fsigss.2009.08.170

- Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability problem. *Human Genomics*, 9:17, 1–12. doi:10.1534/genetics.109.102798
- UCL. (2015). Human Epigenome Project. *University College London*. Retrieved September 10, 2015, from <http://www.epigenome.org/>
- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10(4), 506–13. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1867860>
- Weinberg, W. (1908). Über den nachweis der vererbung beim Menschen., 64, 368–382.
- Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., ... Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron*, 81(1), 103–19. doi:10.1016/j.neuron.2013.10.051
- Wynbrandt, J., & Ludman, M. D. (2009). *The Encyclopedia of Genetic Disorders and Birth Defects*. Infobase Publishing. Retrieved from <http://books.google.com/books?id=kXaMjwltP0oC&pgis=1>
- Zarkower, D. (2001). Establishing sexual dimorphism: conservation amidst diversity? *Nature Reviews Genetics Genet*, 2(3), 175–185. Retrieved from <http://dx.doi.org/10.1038/35056032>
- Zimmer, C. (2014). Seeing X Chromosomes in a New Light - NYTimes.com. *New York Times*. Retrieved March 24, 2014, from http://www.nytimes.com/2014/01/21/science/seeing-x-chromosomes-in-a-new-light.html?_r=1

